

THE
LONDON, EDINBURGH & DUBLIN
PHILOSOPHICAL MAGAZINE
AND
JOURNAL OF SCIENCE

[SEVENTH SERIES—VOL. 32]

I. *A Detailed Analysis of the Gold L Spectrum.*

By E. A. WOODALL, Ph.D.*

[Received December 20, 1940.]

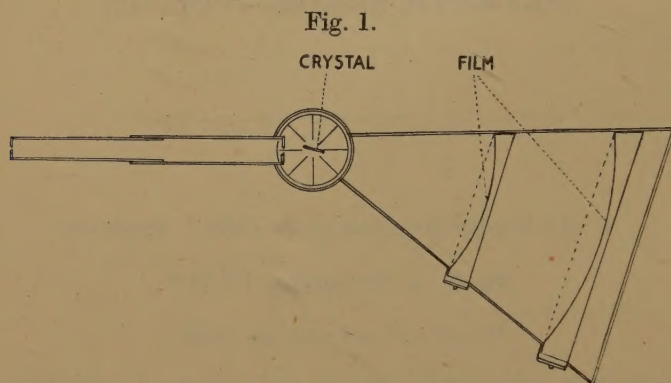
THE present work was undertaken because of the need which was felt for more accurate measurements on the Gold L spectral lines. An examination of the results previously obtained reveals inconsistencies when an attempt is made to calculate the energy level values from them and, in particular, the spin doublet differences vary amongst themselves by as much as 0.45 Rydberg unit.

The spectrograph used, of which a plan is shown in fig. 1, was designed for working in a vacuum at two different distances, and consists of a central cylindrical chamber containing the crystal table, which is graduated in degrees to enable the crystal to be set roughly at the required angle to the incident beam. The table is supported by a spindle which passes through a stuffing box in the base. Attached to the spindle, beneath the spectrograph, is a lever carrying a roller which moves over a slowly rotating cam, so that the crystal is rocked backwards and forwards through a small angle.

* Communicated by the Author.

In rigid connexion with the central chamber are a tube, closed at each end by adjustable slits, and a sector-shaped chamber in which the dark slides containing the photographic films can be inserted at two distances, approximately 24 cm. and 40 cm. from the axis of rotation of the crystal. The tube consists of two parts, one joined rigidly to the spectrograph and the other sliding in and out of this fixed part, so that, when pushed right in as far as possible, the outer slit is 24 cm. from the axis of the crystal table, and when drawn out to a certain mark, this distance is 40 cm.

The outer tube is closed at the end except for a small rectangular opening 1 cm. high and $\frac{1}{2}$ cm. wide. This opening is in a recess in the end of the tube and is closed with a disk of red cellophane and made vacuum-tight with Apiezon black wax. A separate brass cap carrying



Simplified plan of spectrograph.

the symmetrically adjustable slit fits on the end of the tube on two dowel pins, and is also provided with two screws to bring the cap into good contact with the end of the tube. The jaws of the slit are made of a lead-antimony alloy which is easier to work than pure lead, being harder, and therefore enables better edges to be obtained. The width of the slit was measured as accurately as possible with a feeler gauge and varied in this work from about 0.04 to 0.08 mm.

The jaws of the inner slit are separately adjustable, and are attached to a cap which fits on the end of the fixed tube inside the central chamber.

There are two film-holders, one for each distance. Each consists of a brass slide with the back surface ground to the required radius, namely, the distance between the axis of rotation of the crystal and the film. The film is kept pressed to this surface by two strips of blue steel, which were slipped under clips at the top and bottom. A brass shutter runs through a vertical slot in the side of the dark slide, and along grooves

in the top and bottom of the slide, and serves to protect the film from the light, on its way from the dark room to the spectrograph. A rectangular opening in the side of the spectrograph allows the dark slide to be slipped into position and metal springs top and bottom keep it pressed up at the correct distance. On withdrawing the shutter in front of the film, the opening in the spectrograph is closed by a small brass lid, which is then waxed up with Everett's vacuum wax to make the spectrograph vacuum-tight.

The sector-shaped portion of the spectrograph was provided with a lid in addition to the central chamber, in case it should be found necessary to open it for any reason. The lid fits over bolts in the spectrograph and can be screwed down tightly by means of nuts. A special plasticine was used to make the lid vacuum-tight.

With this arrangement the vacuum obtained was no more than about 1 mm. of mercury, and it was necessary to pump the spectrograph continuously to maintain this.

The cam rotates once every 17 mins., and, since the crystal rocks backwards and forwards through $8\frac{1}{2}^{\circ}$ for each revolution of the cam, the speed of rotation of the crystal is 1° per min. The continuous rotation of the cam is obtained with a motor and reduction gearing, which are on a separate smaller table underneath the spectrograph table. By this means, any vibration of the spectrograph which might be caused by the motor is avoided. When the motor was in use, no vibration of the crystal table could be detected when its rotation was observed with a microscope.

The centre part of the spectrograph was supported on a cast iron base by three legs, and the base was provided with three levelling screws. The sector-shaped portion also rested on a support with an adjustable screw.

The crystal-holder is shown in fig. 2. A small brass plate P is attached to the vertical plate of the stand S by three screws carrying springs, which keep the plate P in position. The crystal is fixed to P with wax.

Adjustments of the Crystal.

In the rocking crystal focusing method, it is necessary that the reflecting face of the crystal shall be very accurately parallel to its axis of rotation, and, moreover, shall contain the axis of rotation.

The setting of the reflecting surface parallel to the axis of rotation was made in the usual way, with a piece of parallel-sided glass pressed against the crystal surface and a telescope and illuminated scale at a distance of about 2 metres. The adjustment of the glass plate is believed to be correct to $\frac{1}{2}$ mm. at a distance of 2 metres. The inclination of the crystal face to the axis of rotation should therefore be not greater than 30 seconds of arc.

To set the crystal with the reflecting face containing the axis of rotation, an auxiliary piece of apparatus was used, consisting of two points which were capable of being moved in two horizontal directions at right-angles. The apparatus is shown in fig. 3. A brass stand contains two horizontal

Fig. 2.

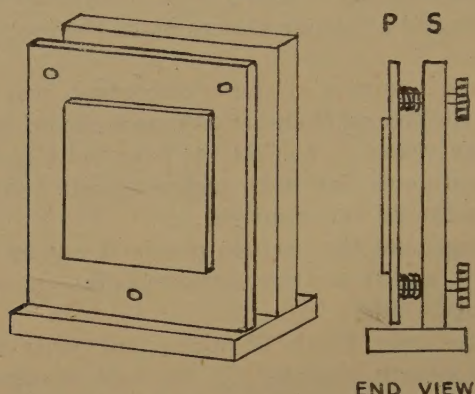
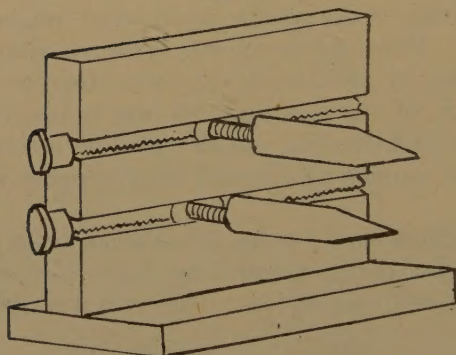


Diagram of crystal-holder.

Fig. 3.



Adjustable points apparatus.

circular grooves, 1 cm. apart. A long fine brass screw runs along each of these grooves, and threaded on each screw is a brass piece carrying a projection at right angles to the screw. Screwed on to this projection is a brass piece with an ivoroid point. The point is thus adjustable in two directions, parallel and perpendicular to the groove respectively.

The apparatus was waxed firmly on the crystal table and, viewing the lower point in a vertical direction by means of a microscope, it was adjusted so that there was no movement of the point in the field of view when the crystal table was rotated. The point was then on the axis

of rotation of the crystal table. If now the crystal were moved up so that it just touched the point, its reflecting face should contain the axis of rotation, for it has already been adjusted to be parallel to this axis. As a test of this first adjustment, the upper point was also set in the axis of rotation. The line joining the two points now is the axis, and if the crystal surface is moved up so that it touches one of the points, it should touch the other without any further adjustment being required. This was found to be the case. When viewed with the microscope, each point was seen to be touching its image in the crystal surface.

With the reflecting face of the crystal touching the points, the crystal-holder was now waxed to the table with paraffin wax, care being taken to keep it accurately in position during the setting of the wax.

Now, having adjusted the slits to be parallel to the points, and therefore to the axis of rotation of the crystal, and after making a rough setting of the crystal to rotate through the range of glancing angles covered by the gold L spectrum, the spectrograph was ready for the taking of photographs.

The X-ray Tube.

The X-ray tube used was a modified form of the Shearer gas tube with air-cooled cathode and water-cooled anticathode. The anticathode could very quickly be removed by breaking a wax seal and a new one waxed into position. The window of the tube was covered with aluminium foil about 0.014 mm. thick. This was put on very quickly by varnishing the surrounding brass with glyptal varnish, and flattening the aluminium foil over the top. A perfectly vacuum-tight window was then obtained by painting round the edges with Everett's wax.

The necessary vacuum was obtained by using a Waran type diffusion pump with mercury, backed by a Cenco Hyvac pump with a three litre aspirator bottle as fore-vacuum.

Preliminary Work.

The shorter distance was used, since such long exposures were required at the longer distance. It was decided to measure the gold L spectrum relative to certain $K\alpha_1$ lines falling in the same range. It was hoped at first that by obtaining these once on a separate photograph and making measurements on them, a calibration equation connecting θ , the glancing angle, and x , the measurement on a spectral line, would be obtained, which could be used in obtaining the glancing angles of the gold L spectral lines from measurements made on any photograph of the gold L spectrum. This was tried out, using as reference lines the $K\alpha_1$ lines of copper, zinc, arsenic, selenium, bromine, strontium, and molybdenum, since these are very well spread over the range of the gold L spectrum and the necessary materials are easy to procure. The lines are also sharp and

relatively easy to measure accurately. For copper $K\alpha_1$ the value 1537.396 XU, recommended by Siegbahn ⁽¹⁾, was used, and for molybdenum, Larsson's ⁽²⁾ value 707.831 XU. Edlén's ⁽³⁾ results for the $K\alpha_1$ lines of the remaining elements were taken and, to avoid any discrepancies which might arise on account of wave-length shift due to chemical combination, the elements were used in the same form, namely zinc dust, arsenious oxide, selenium powder, sodium bromide, strontium oxide. Molybdenum was used in the form of a powder and copper was obtained from the copper backing of the anticathode. The materials were mixed together and the whole made into a paste with water. This was put on a copper anticathode in a thin layer and, when left for a few hours, dried into quite a hard mass. A good photograph was obtained after about 15 hours' exposure, during which it had not been possible to run the tube very hard on account of the sputtering of the material round the sides. This rendered the tube very unsteady in working and the solid particles shot off from the anticathode seemed to act as "getters," absorbing the gas to such an extent that the tube became very hard and refused to pass any current at all.

Some photographs of the gold L spectrum were taken, two of which contained the $K\alpha$ lines of copper and zinc and enabled a comparison with the film containing the reference lines to be made. After making an extensive series of measurements it was found impossible to obtain consistent results, and the original idea had to be abandoned. The reason for this is as follows:—

In order that measurements on different films should be comparable, it is necessary

- (a) to measure always at right-angles to the lines, or at any rate at the same angle to the perpendicular, which should be small since the lines are not perfectly straight;
- (b) to measure the lines at the same height, preferably at the middle, to reduce the effects of asymmetrical broadening;
- (c) to measure the films at the same temperature.

It was found impossible to arrange these conditions, of which (c) was the most difficult, and very large discrepancies were found between measurements made on the same film at different times, so large that they could not be explained as errors in setting on the lines.

Final Work.

It was therefore necessary to obtain the reference lines on each photograph of the gold L spectrum taken. This necessitated changing the anticathode and giving an extra 15 hrs. exposure for the reference lines for each photograph taken, and this, of course, slowed up the work. Also bromine and selenium had to be cut out because the bromine $K\alpha_1$

line falls very close to the gold $L\beta_5$, and the selenium $K\alpha_1$ and α_2 lines are very close to the gold $L\beta_6$ and β_4 . The materials used were therefore molybdenum, strontium, arsenic, zinc and copper, but all these did not appear on every film. The mixture was found to sputter round the tube rather freely, which meant that the tube had to be taken to pieces and cleaned fairly frequently. Later, however, a little water glass was added to the paste and this was very successful. The tube could now be run quite hard and required very little attention.

Measurements.

Some measurements were made on a measuring microscope reading to 0.001 mm. and some on a comparator which enabled measurements to 0.01 mm. to be taken. In each case 20 settings were made on each line, 10 in one direction and 10 in the reverse direction. In addition to the stronger lines measured in this way, there are one or two very faint lines which disappear altogether, or become very difficult to measure under the magnification of the microscope. These were measured with a simple microscope constructed so as to eliminate as far as possible trouble due to parallax. It consisted essentially of two pieces of telescope tubing with very fine cross hairs of silk at one end and a lens at the other. The cross wires were so adjusted that they nearly touched the film and the spectral line was viewed through a small circular aperture some few centimetres away from the lens. In this way no trouble due to parallax was experienced. The simple microscope was attached to the microscope carriage of the comparator so that it traversed the film while the compound microscope traversed the scale. The measurements taken with the simple microscope are distinguished in the tables from the others by the letters SM.

Method of Calculation.

Using the measurements on the reference lines and the known glancing angles, to find the linear relationship $\theta = mx + c$ between the glancing angle θ and the measurement x on a line :—

Film S 8

Measurements on ref. lines.		Glancing angles.
Zn $K\alpha_1$	$x_{Zn} = 45.1285$	$\theta_{Zn} = 13^\circ 40' 28.4''$
As $K\alpha_1$	$x_{As} = 66.1245$	$\theta_{As} = 11^\circ 10' 7.4''$
Sr $K\alpha_1$	$x_{Sr} = 90.2515$	$\theta_{Sr} = 8^\circ 17' 23.0''$
Mo $K\alpha_1$	$x_{Mo} = 103.4915$	$\theta_{Mo} = 6^\circ 42' 35.3''$
	4)304.996	4)39 50 34.1
Mean	$x_m = 76.249$	$\theta_m = 9^\circ 57' 38.5''$

Measurement on ref. lines.	Glancing angles.
$x_m - x_{Zn} = 31.1205$	$\theta_{Zn} - \theta_m = 3^{\circ} 42' 49.9''$
$x_m - x_{As} = 10.1245$	$\theta_{As} - \theta_m = 1^{\circ} 12' 28.9''$
<u>41.245</u>	<u>4 55 18.8</u>
Check :—	
$x_{Sr} - x_m = 14.0025$	$\theta_m - \theta_{Sr} = 1^{\circ} 40' 15.5''$
$x_{Mo} - x_m = 27.2425$	$\theta_m - \theta_{Mo} = 3^{\circ} 15' 3.2''$
<u>41.245</u>	<u>4 55 18.7</u>

$$\therefore \frac{\theta - 9^{\circ} 57' 38.5''}{4^{\circ} 55' 18.8''} = - \frac{x - 76.249}{41.245}$$

$$\text{or } \theta = 9^{\circ} 57' 38.5'' + \frac{17718.8}{41.245} \cdot (76.249 - x)''.$$

$$i. e. \quad \theta = 9^{\circ} 57' 38.5'' + m \cdot (76.249 - x)'', \quad \text{where } \text{Log } m = 2.6330630.$$

$$\begin{aligned} \text{For Au } L\gamma_1, \quad x &= 86.16, \\ \therefore \quad \theta &= 9^{\circ} 57' 38.5'' - m \cdot 9.911'' \\ &= \underline{8^{\circ} 46' 40.7''} \end{aligned}$$

For first order,

$$\begin{aligned} \lambda &= 2d \sin \theta & \text{where } d &= 3029.04 \text{ XU} \\ &= \underline{924.50 \text{ XU}} & \text{Log } 2d &= 3.7823350. \end{aligned}$$

In calculating the following results, the glancing angles of the reference lines at the average temperature of exposure of the lines were used, and the glancing angles of the gold L lines obtained were corrected to 18° C.

Results.

α_1	$L_{III}M_V$	α_2	$L_{III}M_{IV}$	β_1	$L_{II}M_{IV}$
S5	12 8 11.8	S5	12 14 44.7	S5	10 16 57.0
S5	12 8 12.1	S5	12 14 49.2	S5	10 16 51.9
S7	12 8 12.0	S7	12 14 45.5	S7	10 16 55.3
S7	12 8 8.5	S7	12 14 42.2	S7	10 16 52.3
S6	12 8 12.1	S6	12 14 52.6	S6	10 16 59.4
S6	12 8 13.5	S6	12 14 49.2	S6	10 16 52.6
S9	12 8 14.8	S9	12 14 44.0	S9	10 16 55.4
		S9	12 14 45.2	S9	10 16 57.5
Mean	12 8 12.1	S8	12 14 46.9	S8	10 16 49.7
$\lambda = 1273.68$		S8	12 14 53.7	S8	10 16 44.8
$\nu/R = 715.46$				S8	10 16 45.8
		Mean	12 14 47.3		
		$\lambda = 1285.02$		Mean	10 16 52.9
		$\nu/R = 709.15$		$\lambda = 1081.26$	
				$\nu/R = 842.78$	

β_2	$L_{III}N_{IV}$	γ_1	$L_{II}N_{IV}$	γ_2	$L_I N_{II}$
S5	10 9 16.0	S5	8 46 52.0	S5	8 34 9.1
S5	10 9 13.5	S5	8 46 48.6	S7	8 33 58.1
S7	10 9 15.1	S7	8 46 49.5	S7	8 34 11.5
S7	10 9 11.1	S7	8 46 44.2	S9	8 34 7.3
S6	10 9 22.3	S9	8 46 50.8	S9	8 34 17.8
S6	10 9 12.3	S9	8 46 54.4	S8	8 34 1.2
S9	10 9 11.9	S8	8 46 43.1	S8	8 34 5.7
S9	10 9 15.7	S8	8 46 39.6	S8	8 34 7.0
Mean	10 9 14.7	Mean	8 46 47.5	Mean	8 34 7.2
	$\lambda=1068.01$		$\lambda=924.70$		$\lambda=902.62$
	$\nu/R=853.23$		$\nu/R=985.48$		$\nu/R=1009.58$

γ_6	$L_{II}O_{IV}$	γ_3	$L_I N_{III}$	β_4	$L_I M_{II}$
S5	8 33 18.5	S5	8 30 32.4	S7	10 30 18.0
S7	8 33 24.1	S7	8 30 31.0	S7	10 30 5.8
S7	8 33 21.4	S9	8 30 24.3	S9	10 30 14.9
S9	8 33 21.8	S9	8 30 31.9	S9	10 30 14.4
S9	8 33 31.0	S8	8 30 19.5	S8	10 30 8.5
S8	8 33 14.2	S8	8 30 23.3	S8	10 30 11.1
S8	8 33 15.1	S8	8 30 22.2	S8	10 30 11.6
S8	8 33 16.2	S8	8 30 24.2	Mean	10 30 12.0
Mean	8 33 20.3	Mean	8 30 26.1		$\lambda=1104.34$
	$\lambda=901.26$		$\lambda=896.20$		$\nu/R=825.17$
	$\nu/R=1011.10$		$\nu/R=1016.81$		

β_6	$L_{III}N_I$	β_5	$L_{III}O_{IV, V}$	l	$L_{III}M_I$
S7	10 32 46.8	S5	9 51 58.0	S5	13 54 48.6
S7	10 32 48.7	S5	9 52 0.3	S5	13 54 48.3
S9	10 32 48.5	S7	9 52 9.2	S9	13 54 49.1
S9	10 32 47.4	S7	9 52 8.5	S9	13 54 41.0
S8	10 32 41.1	S9	9 52 11.6	S8	13 54 48.6
S8	10 32 46.7	S9	9 52 14.4	S8	13 54 53.6
S8	10 32 43.4	S8	9 52 7.2	S8	13 54 42.6
		S8	9 52 11.8	Mean	13 54 47.4
Mean	10 32 46.1	Mean	9 52 9.6		$\lambda=1456.67$
	$\lambda=1108.79$		$\lambda=1038.31$		$\nu/R=625.58$
	$\nu/R=821.85$		$\nu/R=877.64$		

β_{15}	$L_{III}N_{IV}$	β_3	$L_I M_{III}$	η	$L_{II} M_I$
S7	10 10 13.9	S7	10 7 54.2	S9	11 25 37.8
S7	10 10 10.2	S7	10 7 58.4	S9	11 25 43.9
S9	10 10 20.3	S9	10 7 52.0	S8	11 25 36.9
S9	10 10 20.9	S9	10 7 50.8	S8	11 25 41.6
		S5	10 7 47.6	S8	11 25 33.9
Mean	10 10 16.4	Mean	10 7 52.6	Mean	11 25 38.8
$\lambda=1069.80$		$\lambda=1065.64$		$\lambda=1200.27$	
$\nu/R=851.81$		$\nu/R=855.15$		$\nu/R=759.22$	

β_7	$L_{III}O_I$	β_9	$L_I M_V$	β_{10}	$L_I M_{IV}$
S9	9 57 38.9	S8	9 40 47.1	S8	9 44 38.3
S9	9 57 33.1	S8	9 40 45.0	S8	9 44 42.4
S8	9 57 43.2	S8	9 40 48.8	S8	9 44 41.5
S8 SM	9 57 37.2	S8 SM	9 40 45.3	S8	9 44 42.2
S9	9 57 23.5	S8 SM	9 40 51.9	S8 SM	9 45 0.3
		S8	9 40 51.4	S8 SM	9 45 5.7
Mean	9 57 35.2	Mean	9 40 48.2	Mean	9 44 48.4
$\lambda=1047.79$		$\lambda=1018.64$		$\lambda=1025.60$	
$\nu/R=869.71$		$\nu/R=894.59$		$\nu/R=888.52$	

β_7'	$L_{III}N_{VI, VII}$	$L_{III}O_{II, III}$	γ_4	$L_I O_{II, III}$	
S9	9 56 12.1	S9	9 55 7.2	S8	8 13 0.3
S9	9 56 13.5	S9	9 55 2.3	S8	8 13 3.6
S8	9 56 6.0	S8	9 54 52.1	S8	8 13 0.8
S8	9 56 7.6	S8	9 54 55.1	S9	8 13 1.0
S8	9 56 9.3	S8	9 54 59.2	S9	8 13 11.2
S8 SM	9 56 3.5	S8 SM	9 55 3.3	S8 SM	8 13 10.1
S8	9 56 14.2	S8	9 55 9.1	S8 SM	8 13 12.8
S9	9 56 6.3	S9	9 55 1.9	S9 SM	8 12 59.4
				S9 SM	8 13 3.0
Mean	9 56 9.1	Mean	9 55 1.3	S8	8 13 4.9
$\lambda=1045.30$		$\lambda=1043.33$		Mean	8 13 4.7
$\nu/R=871.78$		$\nu/R=873.42$		$\lambda=865.94$	
				$\nu/R=1052.35$	

γ_5	$L_{II}N_I$	γ_{11}	$L_{II}N_V$	γ_8	$L_{II}O_I$
S9	$\overset{\circ}{9} \overset{\prime}{3} \overset{''}{25.6}$	S8 SM	$\overset{\circ}{8} \overset{\prime}{22} \overset{''}{36.9}$	S9	$\overset{\circ}{8} \overset{\prime}{36} \overset{''}{49.8}$
S9	9 3 26.7	S8 SM	8 22 34.9	S8	8 37 0.2
S8	9 3 28.7	S9 SM	8 22 45.8	S9 SM	8 36 54.9
S8	9 3 31.2	S9 SM	8 22 46.5	S9 SM	8 37 1.3
S8	9 3 27.9			S8 SM	8 36 49.1
		Mean	8 22 41.0	S8 SM	8 36 49.5
Mean	9 3 28.0	$\lambda = 882.69$		Mean	8 36 54.1
$\lambda = 953.73$		$\nu/R = 1032.38$		$\lambda = 907.47$	
$\nu/R = 955.48$				$\nu/R = 1004.19$	

An attempt to
separate γ_4 .

$L_I O_{II}.$	$L_I O_{IV}, \nu.$	s.	$L_{III} M_{III}.$
S8 $\overset{\circ}{8} \overset{\prime}{13} \overset{''}{31.9}$	S8 SM $\overset{\circ}{8} \overset{\prime}{11} \overset{''}{20.9}$	S8 $\overset{\circ}{12} \overset{\prime}{51} \overset{''}{23.2}$	
$\lambda = 866.73$	S8 SM 8 11 4.4	S8 SM 12 51 33.2	
$\nu/R = 1051.38$	S9 SM 8 11 31.4	S8 SM 12 51 15.4	
	S9 SM 8 11 32.8	S9 SM 12 51 31.4	
$L_I O_{III}.$		S9 SM 12 51 52.9	
	Mean 8 11 22.4		
	$\lambda = 862.96$	Mean 12 51 31.2	
S8 $\overset{\circ}{8} \overset{\prime}{12} \overset{''}{49.4}$	$\nu/R = 1055.97$	$\lambda = 1348.21$	
$\lambda = 865.50$		$\nu/R = 675.92$	
$\nu/R = 1052.88$			

Non-diagram Lines.

β_2' .		β_2'' .			
S8	$\overset{\circ}{10} \overset{\prime}{6} \overset{''}{17.2}$	S8	$\overset{\circ}{10} \overset{\prime}{5} \overset{''}{22.3}$	S8	$\overset{\circ}{10} \overset{\prime}{0} \overset{''}{32.1}$
S8	10 6 19.2	S8	10 5 19.8	S9	10 0 38.0
S9	10 6 19.7	S9	10 5 22.6		
S9	10 6 16.4	S9	10 5 20.6	Mean	10 0 35.0
		S5	10 5 20.8	$\lambda=1052.99$	
Mean	10 6 18.1	Mean	10 5 21.2	$\nu/R=865.43$	
$\lambda=1062.91$		$\lambda=1061.26$			
$\nu/R=857.35$		$\nu/R=858.68$			

S8	9 53 51.1	S8	9 48 53.2	S8	9 47 38.1
S8	9 53 56.3	S5	9 48 51.2	S8	9 47 40.6
S9	9 53 47.6	S8 SM	9 48 34.1	S8	9 47 41.4
S9	9 53 48.0	S8 SM	9 48 36.1	S8 SM	9 47 38.6
S5	9 54 3.5			S8 SM	9 47 38.2
S5	9 54 4.6	Mean	9 48 43.6		
S8 SM	9 53 31.7	$\lambda=1032.40$		Mean	9 47 39.4
S8 SM	9 53 48.3	$\nu/R=882.68$		$\lambda=1030.55$	
				$\nu/R=884.26$	
Mean	9 53 51.4				
	$\lambda=1041.31$				
	$\nu/R=875.11$				

The following lines of doubtful origin were also measured :—

S8	9 18 12.6	S8	8 40 51.4	S8	8 44 19.9
S8 SM	9 17 58.0	S8	8 41 7.6		$\lambda=920.42$
S8 SM	9 18 20.1	S8 SM	8 41 10.8		$\nu/R=990.06$
S9 SM	9 18 3.6	S8 SM	8 40 53.0	S9 SM	8 56 28.1
S9 SM	9 17 35.5	S9 SM	8 40 48.0		$\lambda=941.54$
					$\nu/R=967.84$
Mean	9 18 2.0	Mean	8 40 58.2		
	$\lambda=979.07$	$\lambda=914.56$		S9 SM	13 4 41.3
$\nu/R=930.77$		$\nu/R=996.42$			$\lambda=1370.84$
					$\nu/R=664.75$

As a test of the internal consistency of these results the spin doublet ν/R differences were calculated. These are as follows :—

$\beta_1 - \alpha_2$	133.63
$\eta - l$	133.64
$\gamma_5 - \beta_6$	133.63
$\gamma_6 - \beta_5$	133.46
$\gamma_1 - \beta_{15}$	133.67
$\gamma_8 - \beta_7$	134.48

The agreement between the first three is very satisfactory, being much better than is to be expected from the accuracy of the measurements. The doublets $\gamma_6 - \beta_5$ and $\gamma_8 - \beta_7$ do not agree so well. γ_6 is very close to γ_2 and the two lines are very difficult to measure. γ_8 is a very weak line and very difficult to see under the magnification of the compound microscope. However, it was decided to measure these lines again, in the hope that by concentrating on them more accurate results would be

obtained. The opportunity was also taken of making measurements on other weak lines such as γ_{11} , which is very faint, γ_4 , and the weak line on the short wave side of γ_4 . An attempt was made to separate the components of γ_4 , which was obviously double when seen under the microscope; but it is not felt that the results of the measurements on these two close lines are very reliable.

The results of these later measurements are as follows. In the case of γ_{11} and $L_{\text{I}}O_{\text{IV}}$, and the components of γ_4 , the new measurements have been combined with the old ones:—

β_7 .	$L_{III}O_I$.	γ_8 .	$L_{II}O_I$.	$L_{II}N_{VI}$.	
S8	9 57 39.4	S8	8 37 14.2	S8	8 36 27.3
S8	9 57 39.8	S8	8 37 18.7	S8	8 36 18.1
S8	9 57 25.6	S8	8 37 17.4	S9	8 35 55.4
S8	9 57 32.9	S9	8 37 10.4	S9	8 36 14.5
S9	9 57 32.8	S9	8 37 37.8	S9 SM	8 36 39.7
S9	9 57 24.8	S9 SM	8 37 19.7	S9	8 36 6.4
S9	9 57 40.8	S9	8 37 30.1		
S9	9 57 44.5			Mean	8 36 16.9
		Mean	8 37 21.2	$\lambda =$	906.39
Mean	9 57 35.1	$\lambda =$	908.25	$\nu/R =$	1005.40
$\lambda =$	1047.79	$\nu/R =$	1003.34		
$\nu/R =$	869.71				

γ_6 .	$L_{II}O_{IV}$.	γ_2 .	$L_I N_{II}$.	γ_{11} .	$L_I N_V$.
S8	8 33 9.2	S8	8 33 53.9	S8	8 22 20.1
S8	8 33 17.6	S8	8 33 51.6	S8	8 22 44.2
S8	8 33 19.8	S5	8 34 11.1	S8	8 22 36.1
S5	8 33 22.6	S5	8 34 0.4	S8 SM	8 22 36.9
S5	8 33 8.7	S7	8 34 7.3	S8 SM	8 22 34.9
S7	8 33 16.6	S7	8 34 10.5	S9 SM	8 22 46.5
S7	8 33 20.2	S7	8 34 1.5	S9 SM	8 22 45.8
S7	8 33 16.0	S9	8 33 57.7	S9	8 22 49.3
S9	8 33 10.9	S9	8 33 50.9		
S9	8 33 8.1	S9	8 34 2.7	Mean	8 22 39.2
S9	8 33 24.5			$\lambda =$	882.63
S9	8 33 24.9	Mean	8 34 0.8	$\nu/R =$	1032.46
		$\lambda =$	902.44		
Mean	8 33 16.6	$\nu/R =$	1009.81		
$\lambda =$	901.15				
$\nu/R =$	1011.25				

$L_I O_{IV}, \gamma$			Components of γ_4			$L_I O_{III}$		
S8	8 11 12.7		S8	8 13 9.9		S8	8 12 33.3	
S8	8 11 13.9		S8	8 13 31.9		S8	8 12 49.4	
S8	8 10 55.0		S9	8 13 33.4		S9	8 12 50.1	
S8 SM	8 11 20.9							
S8 SM	8 11 4.4		Mean	8 13 25.1		Mean	8 12 44.3	
S9	8 10 38.3		$\lambda = 866.53$			$\lambda = 865.34$		
S9	8 10 28.9		$\nu/R = 1051.65$			$\nu/R = 1053.11$		
S9	8 10 45.4							
S9 SM	8 11 31.4							
S9 SM	8 11 32.8							
			$L_{II} M_{III}$			$L_{II} N_{III}$		
Mean	8 11 4.4		S9 SM	10 42 30.4		S9	8 54 10.8	
$\lambda = 862.44$			$\lambda = 1125.66$			$\lambda = 937.56$		
$\nu/R = 1056.63$			$\nu/R = 809.56$			$\nu/R = 971.97$		

*Non-diagram Lines.*Satellites on α_1

S7	12 5 21.1		S8	10 3 43.9		S9 SM	8 28 15.0	
S5	12 5 12.2		$\lambda = 1058.45$			$\lambda = 892.39$		
			$\nu/R = 860.96$			$\nu/R = 1021.17$		
Mean	12 5 16.7							
$\lambda = 1268.64$								
$\nu/R = 718.32$			S8	9 51 4.4		S8	9 50 1.0	
			$\lambda = 1036.48$			S8	9 49 55.6	
S5	12 6 57.6		$\nu/R = 879.19$					
$\lambda = 1271.54$						Mean	9 49 58.3	
$\nu/R = 716.68$						$\lambda = 1034.56$		
						$\nu/R = 880.84$		

In addition to the above results, several other very weak lines were measured. These, however, are not diagram lines and, since their existence is very doubtful, they are not included here. It is hoped at a later date to be able to establish whether these lines really do exist.

To return to the spin doublet differences, it will be noticed that the ν/R value for β_7 remains unchanged in the new measurements. Inserting the new values of γ_6 and γ_8 we have

$$\gamma_6 - \beta_5 = 1011.25 - 877.64 = 133.61,$$

$$\gamma_8 - \beta_7 = 1003.34 - 869.71 = 133.63.$$

These are in much closer agreement with the other spin doublet values and the mean spin doublet difference is 133.65.

Summary of Results.

Line.	Glancing angle.	Wave-length. XU.	ν/R .
α_1 $L_{III}M_V$	12° 8' 12.1"	1273.68	715.46
α_2 $L_{III}M_{IV}$	12 14 47.3	1285.02	709.15
β_1 $L_{II}M_{IV}$	10 16 52.9	1081.26	842.78
β_2 $L_{III}N_V$	10 9 14.7	1068.01	853.23
γ_1 $L_{II}N_{IV}$	8 46 47.5	924.70	985.48
β_4 $L_I M_{II}$	10 30 12.0	1104.34	825.17
β_6 $L_{III}N_I$	10 32 46.1	1108.79	821.85
β_5 $L_{III}O_{IV, V}$	9 52 7.8	1038.31	877.64
β_3 $L_I M_{III}$	10 7 52.8	1065.64	855.15
β_{15} $L_{III}N_{IV}$	10 10 16.4	1069.80	851.81
γ_2 $L_I N_{II}$	8 34 0.8	902.44	1009.81
γ_6 $L_{II}O_{IV}$	8 33 16.6	901.15	1011.25
γ_3 $L_I N_{III}$	8 30 26.1	896.20	1016.81
l $L_{III}M_I$	13 54 47.4	1456.67	625.58
η $L_{II}M_I$	11 25 38.8	1200.27	759.22
β_7 $L_{III}O_I$	9 57 35.2	1047.79	869.71
γ_4 $L_I O_{II, III}$	8 13 4.7	865.94	1052.35
$\left\{ \begin{array}{l} L_I O_{II} \\ L_I O_{III} \end{array} \right.$	8 13 25.1	866.53	1051.65
$\left\{ \begin{array}{l} L_I O_{II} \\ L_I O_{III} \end{array} \right.$	8 12 44.3	865.34	1053.11
γ_5 $L_{II}N_I$	9 3 28.0	953.73	955.48
γ_8 $L_{II}O_I$	8 37 21.2	908.25	1003.34

Quadrupole lines.

β_9 $L_I M_V$	9° 40' 48.2"	1018.64	894.59
β_{10} $L_I M_{IV}$	9 44 48.4	1025.60	888.52
β'_7 $L_{III}N_{VI, VII}$	9 56 9.1	1045.30	871.78
s $L_{III}M_{III}$	12 51 31.2	1348.21	675.92
γ_{11} $L_I N_V$	8 22 39.2	882.63	1032.46
$L_{II}N_{VI}$	8 36 16.9	906.39	1005.40
$L_I O_{IV, V}$	8 11 4.4	862.44	1056.63
$L_{III}O_{II, III}$	9 55 1.3	1043.33	873.42
$L_{II}N_{III}$	8 54 10.8	937.56	971.97
$L_{II}M_{III}$	10 42 30.4	1125.66	809.56

Non-diagram lines.

β'_2	10° 6' 18.1"	1062.91	857.35
β''_2	10 5 21.2	1061.26	858.68
	9 53 51.4	1041.31	875.11
	9 48 43.6	1032.40	882.68
	9 47 39.4	1030.55	884.26
α''	$\left\{ \begin{array}{l} 12 5 16.7 \\ 12 6 57.6 \end{array} \right.$	$\left\{ \begin{array}{l} 1268.64 \\ 1271.54 \end{array} \right.$	$\left\{ \begin{array}{l} 718.32 \\ 716.68 \end{array} \right.$
	10 0 35.0	1052.99	865.43
	10 3 43.9	1058.45	860.96

Comparison with other Workers.

λ .	Woodall.	Friman ⁽⁴⁾ .	Idei ⁽⁵⁾ .	Claësson ⁽⁶⁾ .	Kaufman ⁽⁷⁾ .	Dauvillier ⁽⁸⁾ .
<i>Dipole lines.</i>						
α_1 $L_{III}M_V$	1273.68	1273.77	—	—	—	1272.8
α_2 $L_{III}M_{IV}$	1285.02	1285.02	—	—	—	—
β_1 $L_{II}M_{IV}$	1081.26	1081.28	—	—	—	1080.9
β_2 $L_{III}N_V$	1068.01	1068.01	—	—	—	1067.5
γ_1 $L_{II}N_{IV}$	924.70	924.61	—	—	—	924.4
l $L_{III}M_I$	1456.67	—	1456.9	—	—	—
η $L_{II}M_I$	1200.27	—	1200.3	—	—	—
β_6 $L_{III}N_I$	1108.79	—	1108.6(3)	—	—	1108.7
β_4 $L_I M_{II}$	1104.34	—	1104.2(2)	—	—	1104.1
β_{15} $L_{III}N_{IV}$	1069.80	—	1069.7	—	—	—
β_3 $L_I M_{III}$	1065.64	—	1065.5(6)	—	—	1065.2
β_7 $L_{III}O_I$	1047.79	—	1047.8	—	—	1047.2
β_5 $L_{III}O_{IV}, v$	1038.31	—	1038.3(0)	—	—	1038.0
γ_2 $L_I N_{II}$	902.44	—	902.4(8)	—	—	902.5
γ_6 $L_{II}O_{IV}$	901.15	—	901.0	—	—	901.2
γ_3 $L_{II}N_{III}$	896.20	—	895.8(1)	—	—	896.1
γ_4 $L_I O_{III}, III$	865.9(4)	—	865.4	—	—	865.8
γ_4 $L_I O_{II}$	866.5(3)	—	—	865.8	—	—
γ_4 $L_I O_{III}$	865.3(4)	—	—	865.5	865.6	—
γ_5 $L_{II}N_I$	953.73	—	953.6	—	—	953.3
γ_F $L_{II}O_I$	908.2(5)	—	907.9	—	—	907.5

Quadrupole lines.

β_9	1018-6(4)	—	1018-8	—	1018-7	1018-2
$L_1 M_V$	1025-6(0)	—	1026-0	—	1025-9	1025-4
β_{10}	1045-3(0)	—	1045-2	—	1045-4	1045-3
β_7	882-6(3)	—	882-8	882-1	882-6	883-0
γ_{11}	1348-2(1)	—	—	—	—	1348-1
δ	1043-3(3)	—	—	—	1043-4	—
$L_{III} M_{III}$	1125-6(6)	—	—	1125-5	—	—
$L_{II} M_{III}$	937-5(6)	—	—	937-4	938-6	—
$L_{II} N_{III}$	906-3(9)	—	—	906-5	906-7	—
$L_{II} N_{VI}$	862-4(4)	—	—	862-0	862-2	—
$L_I O_{IV, V}$	—	—	—	$L_V O_{IV}$ $L_I O_V$	—	—

Non-diagram lines.

$\beta_{2'}$	1062-9(1)	—	1062-8	—	1062-9	1062-3
$\beta_{2''}$	1061-2(6)	—	1061-2	—	1061-3	1061-0
	1058-4(5)	—	—	—	—	—
	1052-9(9)	—	—	—	1052-7	—
	1041-31	—	—	—	—	1041-2
	1036-4(8)	—	—	—	—	1036-0
	1034-5(6)	—	—	—	1035-4	—
	1032-4(0)	—	—	1032-3	1032-3	—
	1030-5(5)	—	—	1031-3	—	—
	$\left\{ \begin{array}{l} 1268-6(4) \\ 1271-5(4) \end{array} \right\}$	—	1268-6	1268-1 1270-4	1029-9	1029-6 1268-9

 α'' *Lines of doubtful origin.*

	892-3(9)	—	—	—	cf. 893-3	—
	914-5(6)	—	—	—	—	920-1
	920-4(2)	—	—	—	—	—
	941-5(4)	—	—	—	—	—
	979-0(7)	—	—	—	—	—
	1370-8(4)	—	—	—	—	—

may be bad measurement of $SbK\alpha_1$ (469.5) in 2nd order.
 may be a measurement of $SnK\alpha_1$ (489.5) in 2nd order.

*Summary of Results (cont.).**Lines of doubtful origin.*

8° 28' 15.0"	892.39	1021.17
8 40 58.2	914.56	996.42
8 44 19.9	920.42	990.06
8 56 28.1	941.54	967.84
9 18 2.0	979.07	930.77
9 49 58.3	1034.56	880.84
9 51 4.4	1036.48	879.19
13 4 41.3	1370.84	664.75

Lines Measured by other Workers, but not in this Work.

λ .	Claësson.	Kaufman.	Dauvillier.
<i>Quadrupole lines.</i>			
$L_I N_{IV}$	883.4	884.0	—
$L_{II} M_V$	—	1072.9	—
$L_{II} O_{III}$	904.9	904.5	—
$L_{III} M_{II}$	—	—	1410.0
$L_{III} N_{II}$	—	1097.7	—
$L_{III} N_{III}$	—	1087.7	—
<i>Non-diagram lines.</i>			
	—	1056.6	—
		(1035.4)	—
	—	899.4	899.0
	—	(893.3)	—
<i>Forbidden line.</i>			
$L_I N_I$	—	911.3	—

*Discussion of Errors.**(a) Non-symmetry of the Lines.*

Measurements have been made to the middle of the lines. This may not coincide with the maximum intensity of the lines. There are various factors which cause an asymmetrical broadening. These are :—

(i.) Penetration into the crystal, when broadening takes place on the short wave side of the line.

(ii.) Height of the slit, which causes broadening on the long wave side.

(iii.) Unequal slit-crystal axis, crystal axis-film distances.

(iv.) Axis of rotation of the crystal not coinciding with the crystal surface. In this case, if the crystal surface, though still parallel to the

axis, is displaced from it through a distance p , the spectral lines will be displaced and also broadened. The amount of displacement will be $2p \cos \theta$, where θ is the glancing angle of the line. Since $\cos \theta$ for the Gold L spectrum varies from 0.990 to 0.970, the displacement of the line can introduce no sensible error in these relative measurements, even assuming that p is as great as 0.01 mm. The broadening (and the displacement) take place towards the short waves if the axis of rotation is outside the crystal, and towards the long waves if the axis of rotation is inside.

Actually, microphotometric recordings of the films showed the lines to be fairly symmetrical, and, since they were so fine, it is believed that no great error can have been introduced by assuming that the region of maximum intensity coincides with the middle of the line.

(b) Asymmetry of the Reference Lines.

In measuring to the middle of the reference lines, it has been assumed that they are symmetrical. If they are not, a slight error will be introduced on this account, provided that the workers whose results have been taken measured to the region of maximum blackening. It is believed, however, that Larsson and Edlén also measured to the middle of the line. If a line is asymmetrical, it is quite probable that, in measuring it, one unconsciously sets on the more intense part.

There is evidence that K lines are not symmetrical. Allison ⁽⁹⁾, for example, has found that a number of $K\alpha_1$ lines are asymmetrical, including those of 29 Cu and 30 Zn. For copper, the ratio of that part of the full width at half maximum lying to the long wave side of the maximum to that on the short wave side was found by him to be 1.4 : 1. For zinc this ratio was more nearly equal to unity. The microphotometric traces of the lines obtained in this work, however, seem to be fairly symmetrical.

(c) Stretching of the Film.

It has been said that films are unsuitable for spectroscopic work because they are liable to stretch during development. For relative measurements this does not affect the results provided that the stretching is uniform. If there were non-uniform stretching of the film, there would be obvious discrepancies between the results of measurements on different films. This was not the case, so that it can be concluded that if the film stretches at all it does so uniformly.

(d) Temperature Effect.

By means of a thermometer placed on the spectrograph, the temperature was taken when starting up and when switching off. In this way

it was possible to make an estimate of the average temperature at which the exposure was taken. Now, if it is possible for the film to creep during these temperature changes (which are frequent, since 50 hrs. exposure was required to bring up the weaker lines, and this occupied several days), it will be impossible to allow accurately for temperature effect. Since, however, the film is held against a brass backing by springs, it is probably safe to assume that there is no slipping between the film and the brass. If the film expands uniformly with the brass, no correction on this account is required. In any case, the temperature effect is very small and, if the results obtained are consistent, this in itself will eliminate any possibility of a large error on this account. It may be added that, if there were a creeping of the film due to temperature changes during the progress of an exposure, the lines obtained would show a broadening due to this. The photographs obtained, however, showed very clear sharp lines, and therefore it can be concluded that there is no slipping between the film and the brass backing while the photograph is being taken.

The correction to the glancing angle, $\Delta\theta = \alpha(18-t) \tan \theta$ radians, where θ is the glancing angle, t is the temperature in $^{\circ}\text{C}.$, and α is the coefficient of expansion of the calcite crystal in a direction perpendicular to the cleavage planes. Putting in the value of α the change (increase) in glancing angle in seconds of arc is $2.145 \tan \theta$ per $^{\circ}\text{C}.$ below $18^{\circ}\text{C}.$

If $\theta = 10^{\circ}$, $\tan \theta = .17633$, and the correction to be made is $0.38''$ per $^{\circ}\text{C}.$ below $18^{\circ}\text{C}.$ It will be seen, therefore, that the temperature effect is small, though not negligible, and is greater for larger glancing angles.

(e) *Errors in the Glancing Angles θ due to Errors in Measurements on the Lines.*

Suppose that there are n reference lines, the measurements on which are $x_1, x_2, x_3, \dots, x_n$, with mean x_m . Also, let the glancing angles of these reference lines be $\theta_1, \theta_2, \theta_3, \dots, \theta_n$, and θ_m their mean. Then the equation representing the linear relationship between θ , the glancing angle of an unknown line, and x , the measurement on it is

$$\theta = \theta_m + \left\{ \frac{\sum_1^n |\theta_m - \theta_r|}{\sum_1^n |x_m - x_r|} \right\} \cdot (x - x_m).$$

This was obtained according to a method due to Awbery⁽¹⁰⁾. In actual practice, in finding the gradient, the values of $\theta_m - \theta_r$ and $x_m - x_r$ where $\theta_m > \theta_r$ and $x_m > x_r$ respectively were found, and the arithmetic checked by calculating these sums when $\theta_m < \theta_r$ and $x_m < x_r$. In what follows it is assumed that the first r measurements are less than x_m .

Now suppose that the errors in measuring the reference lines are $\delta_1, \delta_2, \delta_3, \dots, \delta_n$, and that the error in measuring the unknown line is δ .

Suppose also that the errors in the glancing angles of the reference lines are zero.

Then the error in θ is :—

$$\Delta\theta = \frac{r\theta_m - \overline{\theta_1 + \theta_2 + \dots + \theta_r}}{(rx_m - \overline{x_1 + x_2 + \dots + x_r})^2} \cdot \{(\delta - \delta_m)(rx_m - \overline{x_1 + x_2 + \dots + x_r}) - (x - x_m)(r\delta_m - \overline{\delta_1 + \delta_2 + \dots + \delta_r})\},$$

where δ_m is the error in x_m .

Denoting the term outside the bracket by k ,

$$\begin{aligned} \Delta\theta &= k \cdot \{(\delta - \delta_m)(rx_m - \overline{x_1 + x_2 + \dots + x_r}) - (x - x_m)(r\delta_m - \overline{\delta_1 + \delta_2 + \dots + \delta_r})\} \\ &= k \cdot \{(\delta(rx_m - \overline{x_1 + x_2 + \dots + x_r}) - \delta_m(rx - \overline{x_1 + x_2 + \dots + x_r}) + (\delta_1 + \delta_2 + \dots + \delta_r)(x - x_m))\}. \end{aligned}$$

$$\text{But } x_m = \frac{1}{n} \cdot (x_1 + x_2 + \dots + x_n) \text{ and } \delta_m = \frac{1}{n} \cdot (\delta_1 + \delta_2 + \dots + \delta_n).$$

$$\begin{aligned} \therefore \Delta\theta &= k \cdot \left\{ \delta(rx_m - \overline{x_1 + x_2 + \dots + x_r}) - (\delta_1 + \delta_2 + \dots + \delta_r) \left(\frac{rx}{n} - \frac{x_1 + x_2 + \dots + x_r}{n} - x + x_m \right) \right. \\ &\quad \left. - (\delta_{r+1} + \delta_{r+2} + \dots + \delta_n) \left(\frac{rx}{n} - \frac{x_1 + x_2 + \dots + x_r}{n} \right) \right\} \\ &= k \cdot \left\{ \delta(rx_m - \overline{x_1 + x_2 + \dots + x_r}) - (\delta_1 + \delta_2 + \dots + \delta_r) \left(\frac{x_{r+1} + x_{r+2} + \dots + x_n}{n} - \frac{(n-r)x}{n} \right) \right. \\ &\quad \left. - (\delta_{r+1} + \delta_{r+2} + \dots + \delta_n) \left(\frac{rx}{n} - \frac{x_1 + x_2 + \dots + x_r}{n} \right) \right\}. \end{aligned}$$

$$\text{Now if } \frac{x_{r+1} + x_{r+2} + \dots + x_n}{n} - \frac{(n-r)x}{n} > 0, \text{ i.e. } x < \frac{x_{r+1} + x_{r+2} + \dots + x_n}{n-r},$$

$$\text{and } \frac{rx}{n} - \frac{x_1 + x_2 + \dots + x_r}{n} > 0, \text{ i.e. } x > \frac{x_1 + x_2 + \dots + x_r}{n},$$

the error in θ will be a maximum when

$$\delta_1 = \delta_2 = \delta_3 = \dots = \delta_{r+1} = \dots = \delta_n = \pm 0.005 \text{ mm. and } \delta = \mp 0.01 \text{ mm.},$$

supposing that the maximum error in measuring any reference line is ± 0.005 mm. and the maximum error in measuring the line of unknown glancing angle is 0.01 mm.

It will be seen that if the unknown line lies between the mean of all the lines on one side of the average x_m , and the mean of all the lines on the other side of x_m , the maximum possible error is independent of the position of the line within these limits, and is obtained when the errors in measuring

the reference lines are all in the same direction whilst the error in measuring the unknown line is in the opposite direction. The value of the maximum possible error in this case is:—

$$\pm k \cdot (|\delta + \delta_m|) \cdot \overline{(rx_m - x_1 + x_2 + \dots + x_r)} = \pm \frac{r\theta_m - \overline{\theta_1 + \theta_2 + \dots + \theta_r}}{rx_m - x_1 + x_2 + \dots + x_r} \cdot (|\delta + \delta_m|).$$

If x lies outside these limits, the maximum error in θ will occur when

$$\delta_1 = \delta_2 = \dots = \delta_r = -\delta_{r+1} = -\delta_{r+2} = \dots = \delta_n = 0.005.$$

In this case the value of the maximum possible error is not independent of x , but is greater the further x lies outside these limits.

Referring to the calculation on p. 7, if x lies between

$$\frac{1}{2}(x_{Zn} + x_{As}) = 55.63$$

and

$$\frac{1}{2}(x_{Sr} + x_{Mo}) = 96.87,$$

then the maximum possible error in θ , assuming errors in measuring reference lines not greater than 0.005 mm. and error in measuring x not greater than 0.01 mm., is

$$\pm \frac{17719}{41.245} \cdot 0.015'' = \pm 6.4''.$$

For any line outside these limits, however, a separate calculation must be made.

For Au L α $x = 43.150$, and for the maximum possible error in measuring this line we get

$$\begin{aligned} & \pm \frac{17719}{41.245^2} \cdot (41.245 \cdot \overline{\delta - \delta_m} + 33.099 \cdot \overline{2\delta_m - \delta_1 - \delta_2})'' \\ & = \pm 7.7''. \end{aligned}$$

The maximum variation from the mean in the case of some of the strong lines falls a little outside the limits calculated here, but on the whole the results are satisfactory.

Discussion of the Results.

The measurements made on the strong lines α_2 , β_1 , β_2 agree perfectly with those made by Friman. In the case of α_1 and γ_1 , however, there is a difference of nearly 0.1 XU. If Friman's value for γ_1 is inserted in the doublet difference $\gamma_1 - \beta_{15}$, the figure 133.76 is obtained, whereas the value here is 133.67 and is much closer to the average value obtained here, 133.63(5). This, however, is no proof of the superiority of this measurement since the other component of the doublet, β_{15} , is not an easy line to measure and may possibly be inaccurate.

As for the remaining dipole lines, one or two agree very well with the measurements of Idei, but mostly the results are about 0.1 XU higher.

It may be interesting here to compare the spin doublet differences of Friman's and Idei's results with those in this work.

Friman and Idei.	Woodall.
$\beta_1 - a_2 = 133.62$	$\beta_1 - a_2 = 133.63$
$\eta - l = 133.70$	$\eta - l = 133.64$
$\gamma_5 - \beta_6 = 133.60$	$\gamma_5 - \beta_6 = 133.63$
$\gamma_6 - \beta_5 = 133.76$	$\gamma_6 - \beta_5 = 133.61$
$\gamma_1 - \beta_{15} = 133.70$	$\gamma_1 - \beta_{15} = 133.67$
$\gamma_8 - \beta_7 = 134.05$	$\gamma_8 - \beta_7 = 133.63$

It will be seen that the internal consistency in the former is not nearly so great as in the present work.

The line γ_4 ($L_I O_{II, III}$) is broad and from its appearance would be expected to be double. An attempt was made to measure its components, but this was difficult and the results were not regarded with any great optimism. They do, however, agree moderately well with the results of an attempt to measure the components of $L_{III} O_{II, III}$. These are discussed later. Claësson claims to have resolved γ_4 using microphotometric recordings, and gives a wave-length difference of only 0.3 XU, but it is felt that Claësson's result is too optimistic. Kaufman gives the value 865.6 XU for $L_I O_{III}$, but makes no mention of the other component $L_I O_{II}$. Richtmyer, Barnes and Ramberg⁽¹¹⁾ have also measured the components of γ_4 . They give

$$\begin{array}{lcl} \gamma_4 & L_I O_{III} & 865.10 \\ \gamma_4' & L_I O_{II} & 866.15 \end{array} \text{ as compared with } \begin{array}{lcl} \gamma_4 & L_I O_{III} & 865.3(4) \\ \gamma_4' & L_I O_{II} & 866.5(3) \end{array} \text{ here.}$$

These workers have also measured $\gamma_2, \gamma_8, \beta_3, \beta_{15}$.

J. H. Williams⁽¹²⁾, in some work on the natural widths of L series lines, has also measured β_3 and β_{15} relative to β_2 (1068.01 XU), Friman's value.

Here are the results collected :—

	Woodall.	Richtmyer, Barnes and Ramberg.	Williams.	Idei.
β_3	1065.64	1065.83	1065.61	1065.50
β_{15}	1069.80	1069.7	1069.54	1069.7
γ_2	902.44	902.20	—	902.48
γ_8	908.2(5)	907.9	—	907.9

Although the line γ_2 has a greater intensity than γ_6 , the intensity of blackening of the film is not so great and it is difficult to measure since it is much broader than γ_6 , and in some spectrograms appears more like a broad band on the long wave side of γ_6 .

γ_8 is very weak, and it is hoped to make a more satisfactory set of measurements on this line and its close neighbour $L_{II}N_{VI}$ at a later date.

Energy Level Scheme.

Since the absorption spectrum of Gold was not measured, an entirely independent set of values for the energy levels from L_I downwards cannot be given. However, by starting with a reliable direct measurement of one of the L levels, the remaining levels can be calculated, and, provided that sufficient spectral lines have been measured, a complete set of level values, except for the K level, can be given. This has been done as far as possible and the results are given here.

The most reliable measurements for the L absorption spectrum seem to be those of Sandström. His values are :—

	λ .	ν/R .
L_I	862.2	1056.92
L_{II}	900.9	1011.49
L_{III}	1038.2	877.70

Since L_{III} is used by Siegbahn as a basis for the calculation of the energy levels, it is chosen here so that a direct comparison may be made.

We have then, to start with, $L_{III}=877.70$. Now, using the average spin doublet difference found earlier,

$$\begin{aligned} L_{II}-L_{III} &= 133.63 \\ L_{II} &= 1011.33 \end{aligned}$$

L_I is obtained from values of L_I-L_{III} and L_I-L_{II} calculated in the following manner :—

$L_I-L_{III}=L_I N_V - L_{III} N_V$	$= 1032.46 - 853.23$	
	$= 179.23$	(3)
$L_I O_{IV, V} - L_{III} O_{IV, V}$	$= 1056.63 - 877.64$	
	$= 178.99$	(1)
$L_I O_{II, III} - L_{III} O_{II, III}$	$= 1052.35 - 873.42$	
	$= 178.93$	(1)
$L_I M_{III} - L_{III} M_{III}$	$= 855.15 - 675.92$	
	$= 179.23$	(4)
$L_I M_V - L_{III} M_V$	$= 894.59 - 715.46$	
	$= 179.13$	(4)
$L_I M_{IV} - L_{III} M_{IV}$	$= 888.52 - 709.15$	
	$= 179.37$	(4)
Weighted Mean	$L_I - L_{III} = 179.21$	

$$\begin{aligned}
L_I - L_{II} = L_I M_{IV} - L_{II} M_{IV} &= 888.52 - 842.78 \\
&= 45.74 & (4) \\
L_I M_{III} - L_{II} M_{III} &= 855.15 - 809.56 \\
&= 45.59 & (4) \\
L_I O_{IV, V} - L_{II} O_{IV} &= 1056.63 - 1011.25 \\
&= 45.38 & (1) \\
\text{Weighted Mean} \quad L_I - L_{II} &= 45.63
\end{aligned}$$

The values for the differences $L_I - L_{III}$ and $L_I - L_{II}$ do not agree so well as the spin doublet differences. This is to be expected, since each pair of lines used contains a quadrupole line and these, being weaker than the dipole lines, are more difficult to measure accurately. To obtain the best value of L_I from these differences they have been weighted according to the figures in the brackets. The weighted mean gives for $L_I - L_{III}$ the value 179.21 and for $L_I - L_{II}$ the value 45.63. These give for L_I 1056.91 and 1056.96 respectively; and, weighting these in the ratio 2 : 1, since the first comes from six differences and the second from three, we get for the best value

$$\underline{L_I = 1056.94}$$

Now

$$\begin{aligned}
L_{III} &= 877.70 \\
L_{II} &= 1011.33 \\
L_I &= 1056.94
\end{aligned}$$

$\eta = L_{II} - M_I$	M_I	$= 1011.33 - 759.22$	} $M_I = 252.11$
		$= 252.11$	
$l = L_{III} - M_I$	M_I	$= 877.70 - 625.58$	
		$= 252.12$	
$\beta_4 = L_I - M_{II}$	M_{II}	$= 1056.94 - 825.17$	} $M_{II} = 201.78$
		$= 231.77$	
$\beta_3 = L_I - M_{III}$	M_{III}	$= 1056.94 - 855.15$	
		$= 201.79$	} $M_{III} = 201.78$
$L_{II} - M_{III}$	M_{III}	$= 1011.33 - 809.56$	
		$= 201.77$	
$s = L_{III} - M_{III}$	M_{III}	$= 877.70 - 675.92$	} $M_{III} = 201.78$
		$= 201.78$	
$\beta_{10} = L_I - M_{IV}$	M_{IV}	$= 1056.94 - 888.52$	
		$= 168.42$	} (1) $M_{IV} = 168.54$
$\beta_1 = L_{II} - M_{IV}$	M_{IV}	$= 1011.33 - 842.78$	
		$= 168.55$	
$\alpha_2 = L_{III} - M_{IV}$	M_{IV}	$= 877.70 - 709.15$	} (5) $M_{IV} = 168.54$
		$= 168.55$	
$\beta_9 = L_I - M_V$	M_V	$= 1056.94 - 894.59$	
		$= 162.35$	} (1) $M_V = 162.26$
$\alpha_1 = L_{III} - M_V$	M_V	$= 877.70 - 715.46$	
		$= 162.24$	

$\gamma_5 = L_{II} - N_I$	N_I	$= 1011.33 - 955.48$	}	$N_I = 55.85$
		$= 55.85$		
$\beta_6 = L_{III} - N_I$	N_I	$= 877.70 - 821.85$		
		$= 55.85$		
$\gamma_2 = L_I - N_{II}$	N_{II}	$= 1056.94 - 1009.81$	}	$N_{II} = 47.13$
		$= 47.13$		
$\gamma_3 = L_I - N_{III}$	N_{III}	$= 1056.94 - 1016.81$		
		$= 40.13$		
$\gamma_1 = L_{II} - N_{IV}$	N_{IV}	$= 1011.33 - 985.48$	}	$N_{IV} = 25.85$
		$= 25.85$		
$\beta_{15} = L_{III} - N_{IV}$	N_{IV}	$= 877.70 - 851.81$		
		$= 25.89$		
$\gamma_{11} = L_I - N_V$	N_V	$= 1056.94 - 1032.46$	}	$N_V = 24.47(5)$
		$= 24.48$		
$\beta_2 = L_{III} - N_V$	N_V	$= 877.70 - 853.23$		
		$= 24.47$		
$L_{II} - N_{VI}$	N_{VI}	$= 1011.33 - 1005.40$		
		$= 5.93$		
$\beta_7' = L_{III} - N_{VI, VII}$	$N_{VI, VII}$	$= 877.70 - 871.78$		
		$= 5.92$		
$\gamma_8 = L_{II} - O_I$	O_I	$= 1011.33 - 1003.34$	}	$O_I = 7.99$
		$= 7.99$		
$\beta_7 = L_{III} - O_I$	O_I	$= 877.70 - 869.71$		
		$= 7.99$		
$\gamma_4 = L_I - O_{II, III}$	$O_{II, III}$	$= 1056.94 - 1052.35$	(2)	Weighted Mean $O_{II, III} = 4.49$
		$= 4.59$		
$L_{III} - O_{II, III}$	$O_{II, III}$	$= 877.70 - 873.42$		
		$= 4.28$	(1)	
$\gamma_6 = L_{II} - O_{IV}$	O_{IV}	$= 1011.33 - 1011.25$	}	$O_{IV} = 0.08$
		$= 0.08$		
$\beta_5 = L_{III} - O_{IV, V}$	$O_{IV, V}$	$= 877.70 - 877.64$		
		$= 0.06$		$O_V = 0.04$

The lines β_7 , β_7' and $L_{III}O_{II, III}$ lie very close together and are difficult to measure, especially as they do not appear to be single. An attempt was made to separate β_7' and $L_{III}O_{II, III}$ with the following results:—

$L_{III}O_{II, III}$:

	$L_{III}O_{III}$	$L_{III}O_{II}$
S8	9 54 44.4	9 55 23.5
S9	9 54 48.4	9 55 33.5
	9 54 46.4	9 55 28.5
	$\lambda = 1042.90$	$\lambda = 1044.12$
	$\nu/R = 873.80$	$\nu/R = 872.78$

β_7' :

	$L_{III}N_{VII}$	$L_{III}N_{VI}$
S8	$\overset{\circ}{9} \overset{\circ}{56} \overset{\circ}{2.0}$	$\overset{\circ}{9} \overset{\circ}{56} \overset{\circ}{26.5}$
	$\lambda=1045.09$	$\lambda=1045.80$
	$\nu/R=871.97$	$\nu/R=871.38$

We are now in a position to calculate the ν/R values of the O_{II} , O_{III} , N_V , and N_{VII} levels. Using also the components of γ_4 we have :—

$$\begin{array}{llll}
 L_{III}-N_{VI} = 871.38 & N_{VI} = 877.70- & 871.38 & \\
 & = & 6.32 & \\
 L_{III}-N_{VII} = 871.97 & N_{VII} = 877.70- & 871.97 & \\
 & = & 5.73 & \\
 L_{III}-O_{II} = 872.78 & O_{II} = 877.70- & 872.78 & \\
 & = & 4.92 & \\
 L_I-O_{II} = 1051.65 & O_{II} = 1056.94- & 1051.65 & \\
 & = & 5.29 & \\
 L_{III}-O_{III} = 873.80 & O_{III} = 877.70- & 873.80 & \\
 & = & 3.90 & \\
 L_I-O_{III} = 1053.11 & O_{III} = 1056.94- & 1053.11 & \\
 & = & 3.83 &
 \end{array}
 \left. \begin{array}{l} \\ \\ \\ \\ \end{array} \right\} \begin{array}{l} O_{II}=5.10 \\ \\ \\ O_{III}=3.87 \end{array}$$

The O_{III} values do not compare badly with each other and the O_{II} are sufficiently close to give the order of magnitude for this level.

The two values of N_{VI} obtained from $L_{II}N_{VI}$ and $L_{III}N_{VI}$ differ also by as much as 0.4 unit. The doublet difference $L_{II}-N_{VI}-(L_{III}-N_{VI})$ is $1005.40-871.38=134.02$. It is felt that the measurement of $L_{II}N_{VI}$ is not very reliable, and Kaufman's value for this line ($\lambda=906.7$, $\nu/R=1005.0$) is in much better agreement, giving $N_{VI}=1011.33-1005.0=6.33$ as compared with 6.32 from $L_{III}N_{VI}$ obtained above. There is, however, no reason for assuming that $L_{III}N_{VI}$ is more accurate than $L_{II}N_{VI}$.

The line $L_I O_{IV, V}$, giving $O_{IV, V}=0.31$ does not fit the scheme very well, and the value obtained in a previous trial set of measurements, $\lambda=862.17$, $\nu/R=1056.94$, giving $O_{IV, V}=0$, seems to be much nearer the truth.

$L_{II}N_{III}$, a very weak line, was measured only once and is inaccurate.

The doublet s ($L_{III}M_{III}$), $L_{II}M_{III}$ gives a value for $L_{II}-L_{III}$ equal to $809.56-675.92=133.64$, which agrees perfectly with the spin doublet difference 133.63(5). Since, however, $L_{II}M_{III}$ is very faint and was measured only once, with the simple microscope, this agreement is accidental. It is believed though that the results given here for this line and for $L_{III}M_{III}$ are very accurate, since they give values for M_{III} which compare very well with that obtained using β_3 ($L_I M_{III}$).

We now give a table comparing the energy level values calculated here with those given by Siegbahn. The differences are shown in the third column. It will be seen that the agreement is fairly good though not perfect. There is, however, no outstanding difference in any of the

energy levels. In general the levels of lower energy have numerically lower values than Siegbahn's, whereas the outer levels have greater numerical values. The difference is about one- or two-tenths of a Rydberg unit, but in the case of the $O_{\text{II, III}}$ level is as large as 0.4 unit.

Final Table of Energy Levels compared with Siegbahn's Values.

	$\nu/R.$	Siegbahn ⁽¹³⁾ $\nu/R.$	$\Delta\nu/R.$
L_I	1056.94	1057.1	-0.2
L_{II}	1011.33	1011.4	-0.1
L_{III}	(877.70)	877.70	
M_I	252.11	252.2	-0.1
M_{II}	231.77	231.8	—
M_{III}	201.78	201.9	-0.1
M_{IV}	168.54	168.6	-0.1
M_V	162.26	162.2	+0.1
N_I	55.85	55.8	—
N_{II}	47.13	47.3	-0.2
N_{III}	40.13	39.9	+0.2
N_{IV}	25.87	25.8	+0.1
N_V	24.47	24.5	—
N_{VI}	6.3(2)	$N_{VI, VII} = 5.9(2)$	+0.1
N_{VII}	5.7(3)		-0.1
O_I	7.99	7.8	+0.2
O_{II}	5.1(0)	$O_{II, III} = 4.4(9)$	+0.4
O_{III}	3.8(7)		
O_{IV}	0.08	$O_{IV, V} = 0.06$	+0.1
O_V	0.04		

Non-diagram Lines.

The strongest of these lines are the spark lines β_2' and β_2'' . Of these, β_2'' was found to be stronger than β_2' since it could be measured on films where β_2' did not appear. This is contrary to the findings of other workers. The values given here agree to 0.1 XU with those given by Kaufman, who believes his measurements of these strong satellites to be superior to those reported before. They are, to 0.1 XU, in agreement with those reported by Richtmyer and Kaufman⁽¹⁴⁾, who give for $L\beta_2'$, $\lambda=1062.86$ and for $L\beta_2''$, $\lambda=1061.28$, and may be the same results expressed to 0.1 XU.

Another strong satellite is α'' on the short wave side of α_1 . This appears more like a band and, according to Richtmyer, who used the ionisation method in conjunction with a double-crystal spectrometer, consists of four components. Only two have been measured by photographic methods, however, until some work by Cauchois⁽¹⁵⁾ in 1937, and two have

been measured here, but with great difficulty. The one of lower frequency, measured on one film only, is not considered to be accurate. Cauchois has given seven satellites for $L\alpha_1$ of gold as follows :—

	λ .	ν/R .
α^x	1268.63	718.31
α_{II}^x	1269.2	717.99
α^{ix}	1270.35	717.34 Sharpest and most
α'	1271.46	716.71 [intense.
α^y	1272.15	716.32
α_{II}^z	1274.6	714.95
α_{III}^z	1275.5	714.44

The wave-lengths obtained by direct measurement here, namely 1268.6(4) and 1271.5(4), actually agree quite well with two of Cauchois' lines, but this is regarded as accidental. Richtmyer and Kaufman have measured photographically only two $L\alpha_1$ satellites. These are, using their notation, $L\alpha^{ix}$, $\lambda=1270.54$ and $L\alpha^x$, $\lambda=1268.63$, both measured relative to Friman's value for $L\alpha_1$.

The line 1058.4 was measured only on the film S8. It has not been measured by any other worker and may possibly be spurious since, when concentrating on weak lines, it seems to be possible to imagine the existence of a line and to set on an accidental grouping of the grains.

The line 1053.0 was also very weak, but was measured on S8 and S9, so that there seems to be no doubt about its existence. It is probably the line 1052.7 measured by Kaufman and called β_2^{iv} by him.

A line, 1041.3, was measured on S5, S8, and S9. It has previously been measured only by Dauvillier, who gave 1041.2, and its origin is unknown. Besides β_2' and β_2'' , Kaufmann has measured five lines which he identifies as satellites of β_2 , and although these range from 1029.9 to 1056.6 and include lines which must be fainter than this one, this line was not measured by him.

Dauvillier also measured a line 1036.0, and Kaufman gives one of wave-length 1035.4, which he calls β_2^v . Two lines, 1034.6 and 1036.5, were measured here and, since in general Dauvillier's values are lower than those found in this work, the latter line may be his 1036.0.

Cläesson has measured two lines, 1032.3 and 1031.3, Kaufman the two lines 1032.3 and 1029.9, and Dauvillier 1031.6 and 1029.6. Here we have measured 1032.4 and 1030.6. It is just possible that three lines exist and that 1030.6 is a mean measurement of 1029.9 and 1031.3.

Corresponding to the very weak line 920.4 measured here, Dauvillier mentions a very faint line 920.1.

The line 892.4 may correspond to one measured by Kaufman, for which he gives 893.3.

941.5 is probably a bad measurement of Sb $K\alpha_1$ (469.5) in the second order, excited in the material of the lead-antimony slit. 979.1 is believed to be a measurement of Sn $K\alpha_1$ (489.5) in the second order, excited in the soldered join of the anticathode.

It is doubted very much whether 914.6 is a measurement of a gold line. It will be noticed that this figure is roughly twice 457, whereas another line measured, 1370.8, is roughly three times 457. The latter was measured with a simple microscope and it is much fainter than the former, suggesting that the two lines are measurements of the same line in second and third order respectively. No record has been found, however, of a line having a wave-length approximately 457 XU.

Non-diagram lines are difficult to identify because they do not fit into the energy level scheme. The method of identification seems to be to plot $\sqrt{\nu/R}$ against atomic number and to compare the slope of the line obtained with that of the supposed parent line. This method cannot, of course, be applied when only one element has been examined and, pending the analysis of other spectra, the existence of those lines which have not been measured by other workers must remain doubtful.

Summary.

A simple spectrograph has been built to enable relative measurements of X-ray spectra to be made over a limited range of glancing angles. Using a calcite crystal, the X-ray spectrum of gold has been examined thoroughly with a view to checking the X-ray energy level values. In all, thirty diagram lines of gold have been measured in addition to several spark lines. The measurements have been made relative to the $K\alpha_1$ lines of certain lighter elements, which fall within the range of the gold L spectrum. The advantages of using such reference lines instead of the strong L lines of gold are :—

(1) These K lines are very sharp, having natural widths much narrower than the L lines. They can therefore be measured more precisely and it is believed that the wave-lengths of the K lines are more reliable than those previously obtained for the gold L lines.

(2) The K lines are obtained separately on the film and therefore do not become overexposed as do the strong L lines when the weaker lines are obtained.

(3) Since all measurements on the L spectrum of gold since 1926, when Friman measured the strong lines α_1 , α_2 , β_1 , β_2 , and γ_1 , have been made relative to these measurements of Friman, usually the value for β_1 ,

it seems to be desirable to have a completely independent set of values in case of unknown errors in Friman's work, especially as Friman took only two plates for each line.

When relative measurements are being made, if the distance between the slit and the plate is not known, then at least two reference lines are required on the plate. The method has been used by other workers of calibrating the spectrogram in seconds of arc per millimetre, by measuring the distances between pairs of reference lines, and using the distance between the unknown spectral line and the nearest reference line in order to calculate the required glancing angle. It is felt that the method used here of obtaining the best linear relationship to fit the measurements on the reference lines and their known glancing angles is definitely superior to this, especially as the glancing angles of the reference lines are themselves subject to errors which will not be the same for each.

The whole of the L spectrum of gold does not seem to have been measured by one person since the work of Coster and Dauvillier respectively in 1921 and 1922. Since then, great improvements have been made in spectroscopic technique and a much greater precision has been obtained. Friman made some precision measurements on the five strongest lines of gold, and Idei, using the same apparatus, measured the remaining dipole lines and one or two quadrupole lines relative to Friman's value for β_1 . If the measurements of Friman and Idei are subject to systematic errors, these will, most probably, be different for the two workers, and the collected results will therefore not be consistent among themselves. Hence the advantage of having the whole spectrum measured by one worker.

In the present work, the internal consistency of the measurements on the dipole lines, as judged by the constancy of the spin doublet differences, is good, and the mean value for this difference, $L_{II} - L_{III} = 133.63(5)$ is believed to be more accurate than that given by Siegbahn, namely 133.68. This last-named figure is the average of values which differ from one another by as much as 0.23 Rydberg unit.

The values for the differences $L_I - L_{III}$ and $L_I - L_{II}$ do not agree so well; but, since these involve quadrupole lines which are much weaker on the whole and therefore more difficult to measure than the dipole lines, this is to be expected. Also, the lines with the broader L_I level as the initial level are more diffuse than the others, and this decreases the accuracy obtainable in measuring them.

Using the value 877.70 obtained by Sandström for L_{III} , a table of energy levels for gold has been calculated from L_I onwards, and it is believed that the values for L_{IV} , M_I , M_{III} , M'_{IV} , N_I , N_{IV} , and N_V , which depend on the spin doublet difference, are accurate.

Acknowledgments.

I wish to thank Prof. H. R. Robinson, of Queen Mary College, for his interest, encouragement and helpful advice during the progress of this work. My sincere thanks are also due to Prof. H. Dingle, of Imperial College, for his kindness in putting at my disposal a measuring microscope and microphotometer.

References.

- (1) M. Siegbahn, "The Methods of Precision Measurements of X-ray Wavelengths," *Ark. Mat. Astr. o. Fys* (A), xxi. Nr. 21 (1929).
- (2) A. Larsson, "Precision Measurements of the K-series of Molybdenum and Iron," *Phil. Mag.* iii. p. 1136 (1927).
- (3) B. Edlén, "Präzisionsmessungen in der K-Serie der Elemente 30 Zn bis 44 Rn," *Zeits. für. Phys.* lii. p. 364 (1929).
- (4) E. Friman, "Präzisionsmessungen in der L-Reihe der Elemente Wolfram bis Uran," *Zeits. für Physik*, xxxix. p. 813 (1926).
- (5) S. Idei, "The Precise Measurements of the L-group of the X-rays in the Heavy Elements," *Sci. Rep. Tôhoku Imp. Univ.* xix. p. 559 (1930).
- (6) H. Claësson, "Schwache Linien des L-Spektrums bei 79 Au und 74 W," *Zeits. für Phys.* lxxxv. p. 231 (1933).
- (7) S. Kaufman, "Faint Lines in the L spectrum of the Elements 73 Ta to 90 Th." Pt. I. *Phys. Rev.* xlv. p. 385 (1934); Pt. II. *Phys. Rev.* xlv. p. 613 (1934).
- (8) M. A. Dauvillier, "Analyse de la Structure Electronique des Eléments," *J. de Physique*, iii. p. 221 (1922).
- (9) S. K. Allison, "The Natural Widths of the $K\alpha$ X-ray Doublet from 26 Fe to 47 Ag," *Phys. Rev.* xlv. p. 63 (1933).
- (10) J. H. Awbery, *Proc. Phys. Soc.* xli. p. 384.
- (11) F. K. Richtmyer, S. W. Barnes, and E. Ramberg, "The Widths of the L-Series Lines and of the Energy Levels of Gold (79)," *Phys. Rev.* xlv. p. 843 (1934).
- (12) J. H. Williams, "The Natural Widths of the L-Series Lines in the X-ray Spectra of the Elements 74 to 83," *Phys. Rev.* xlv. p. 71 (1934).
- (13) M. Siegbahn, 'Spektroskopie der Röntgenstrahlen' (1930).
- (14) Richtmyer and Kaufman, "X-ray Satellites of High Atomic Number Elements," *Phys. Rev.* xlv. p. 605 (1933).
- (15) Y. Cauchois, "Etude de Satellites $L\alpha$ pour quelques Eléments Lourds et Terres Rares," *Comptes Rendus*, ccv. p. 519 (1937).

The compatibility equations are thereby satisfied. The shear stresses are then given by

$$\widehat{r\theta} = -\frac{1}{r^2} \frac{\partial \phi}{\partial z} \quad \dots \quad (2.2)$$

$$\widehat{\theta z} = \frac{1}{r^2} \frac{\partial \phi}{\partial r} \quad \dots \quad (2.3)$$

$$\widehat{zr} = 0, \quad \dots \quad (2.4)$$

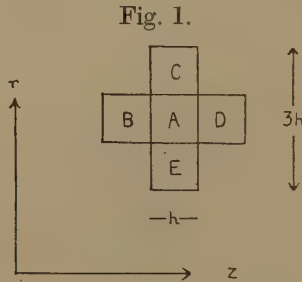
and the applied torque M by

$$M = 2\pi[\phi]_b^a, \quad \dots \quad (2.5)$$

where a and b are the external and internal radii respectively for the section over which the torque is applied.

III. Finite Difference Formulæ and Equations.

The problem is therefore reduced to the numerical determination of the function ϕ over the two-dimensional region defined by the plane



of the meridional section. The region is divided up into square cells of side h in the manner indicated by figs. 2, 5, and 7 for the various sections treated.

Consider the five adjacent cells in fig. 1. If the point A is denoted by coordinates $(0, 0)$, then, B, C, D, E have coordinates $(\bar{1}, 0), (0, 1), (1, 0), (0, \bar{1})$ respectively, the unit coordinate being the length h . The value of ϕ at the centre of a square where the coordinate values are p, q will be denoted by $[p, q]$.

Then the various partial derivations of ϕ which are required are given by the following relations :

$$\frac{\partial \phi}{\partial r} = \frac{[0, 1] - [0, \bar{1}]}{2h} \quad \text{at } A, \quad \dots \quad (3.1)$$

$$\frac{\partial \phi}{\partial z} = \frac{[1, 0] - [\bar{1}, 0]}{2h} \quad \text{at } A, \quad \dots \quad (3.2)$$

$$\frac{\partial^2 \phi}{\partial r^2} = \frac{[0, 1] + [0, \bar{1}] - 2[0, 0]}{h^2} \text{ at A, } \quad \dots \quad (3.3)$$

$$\frac{\partial^2 \phi}{\partial z^2} = \frac{[1, 0] + [\bar{1}, 0] - 2[0, 0]}{h^2} \text{ at A. } \quad \dots \quad (3.4)$$

We may then transform equation (2.1) into an equation relating the ϕ values at the centres of five adjacent squares, obtaining

$$[0, 0] = \frac{(2p-3)[0, 1] + (2p+3)[0, \bar{1}] + 2p\{[1, 0] + [\bar{1}, 0]\}}{8p}, \quad \dots \quad (3.5)$$

where $r = ph$.

IV. Sequence Method of Approximating to the Correct ϕ Values.

The region for which the ϕ values (referred to on some occasions as the "body" values) are required is divided into square cells of side h . The boundary of the region passes through the centres of the squares on the boundaries of the assemblage of cells as shown in figs. 2, 5, and 7. By some method such as that described in Section V., or, more simply, by guesswork, a set of body values is written down which satisfies the given boundary conditions of the problem. These values, however, will not be such as to satisfy equation (3.5) for cells within the region for which the solution is required. To obtain a second set of body values which approximate more closely to a solution of (3.5) the following procedure is adopted.

The value of ϕ , viz. $[p, q]$, at the centre of any square (p, q) is replaced by $[p, q]'$, where

$$8p[p, q]' = \frac{(2p-3)[p, q+1] + (2p+3)[p, q-1] + 2p\{[p+1, q] + [p-1, q]\}}{\dots} \quad (4.1)$$

Every body value is then replaced by the value obtained from the above equation, and the whole process is then repeated, $[p, q]'$ being replaced by $[p, q]''$, where

$$8p[p, q]'' = (2p-3)[p, q+1]' + (2p+3)[p, q-1]' + 2p\{[p+1, q]' + [p-1, q]'\}, \quad \dots \quad (4.2)$$

for every square in the region except the boundary ones, which retain their original values, in order to satisfy the boundary conditions. The whole procedure is then continued until the difference between the body values for any square, between two successive stages of approximation, does not exceed a certain desired percentage of the final value. In view of the fact that the first-order partial derivatives of ϕ determine the stress values, and that equations (3.1) and (3.2) show that these derivatives involve the differences of adjacent ϕ values, then greater accuracy

squares, but for the mid-points of their edges. The values corresponding with the centres of the squares were obtained by graphical interpolation, and the shear stresses evaluated according to equations (2.2) and (2.3).

Fig. 3.
E F G H I J

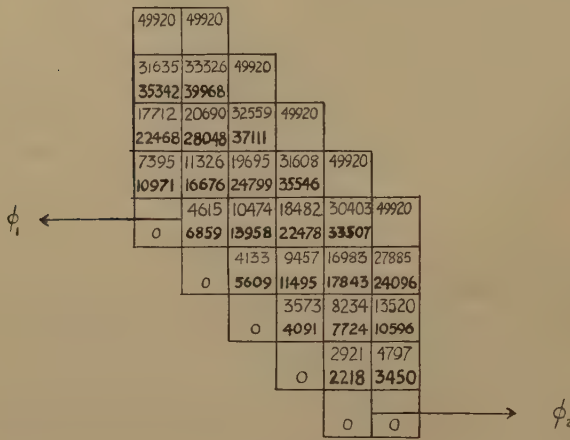
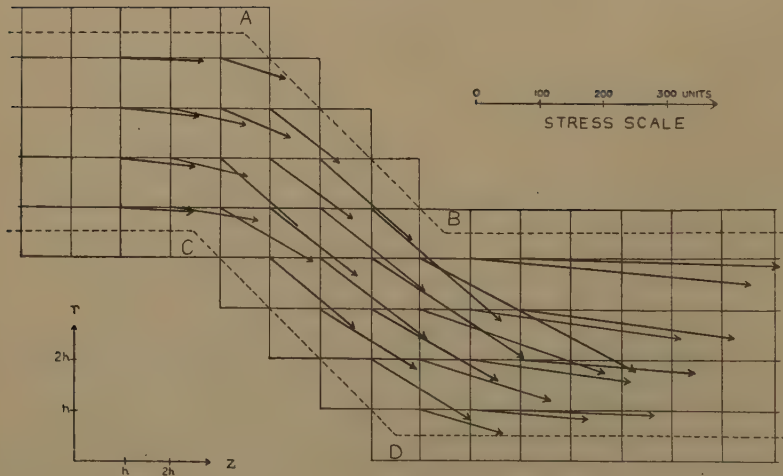


Fig. 4.



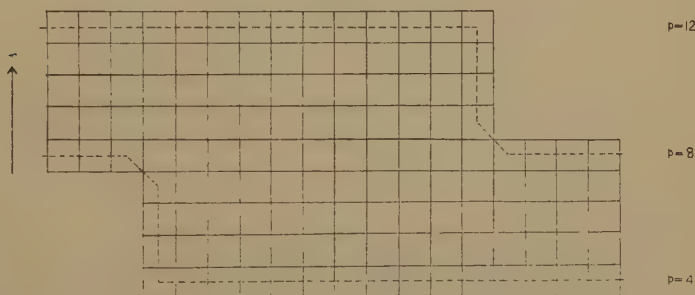
The stresses are shown for the immediate neighbourhood of the neck in fig. 4, in which the arrows represent the resultant shear stress in magnitude and direction at the points in question.

These results show that over the neck, for a given z value, there is little variation of resultant stress with r , except at its extremities in the neighbourhood of the points A, B, C, D. At the "wide" end of the neck the greatest stress occurs near the point C on the inner boundary, while at the "narrow" end it occurs near B on the outer boundary. For regions well removed from the neck the linear variation of $\hat{\theta}z$ with r for a given z is almost exactly observed as the forms of ϕ_1 and ϕ_2 require. In such regions the resultant stress for a given r value tends to be independent of z , which is precisely what would be anticipated in view of the closer approximation of the ϕ values to those given by ϕ_1 and ϕ_2 at large distances from the neck. Near the neck it is observed that the resultant shear increases with z for a given r . It would appear, therefore, that the failure of such a shaft under torsion is most likely to occur in the neighbourhood of the point B.

VI. The Torsion of a Hollow Shaft with Neck Perpendicular to its Axis.

The meridian section of this shaft is illustrated by the broken line in fig. 5. It will be noted that the concave junctions of the shaft and neck

Fig. 5.



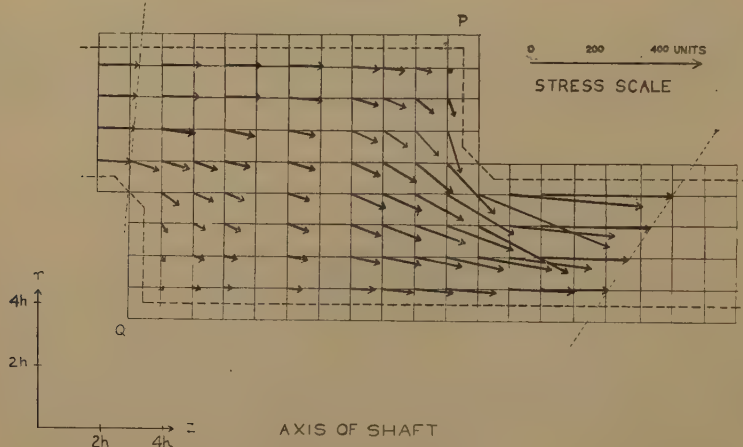
are chamfered at 45° . This is necessary in view of the following considerations. If a finite portion of the shaft is considered, over which boundary values for ϕ have been assigned, then there are, say, n unknown ϕ values to be determined within this region. There will be n equations of the type (3.5) centred on the relevant squares, if the corners are chamfered. If, however, they are not chamfered, the corner squares have ϕ values assigned to them from the boundary conditions, and although there are still n equations, there are only $(n-2)$ unknown ϕ values to be determined. In the former case the n equations suffice to determine the n unknowns, which are then determinate. In the latter case only those values which best "fit" the n equations will be derived from the sequence process. These considerations are valid however large n is,

and though the shaft is considered to be of infinite extent, yet at any stage only a finite number of squares are dealt with, and to make the problem exactly determinate at each stage concave corners must be considered as chamfered*.

The initial set of ϕ values was constructed in a manner similar to that described in Section V. Three functions ϕ_1 , ϕ_2 , and ϕ_3 , each of the type $\phi = Ar^4 + B$, were constructed: one for the region to the left of the neck, one for the centre of the neck, and one for the region to the right of the neck. The other columns in the neck were assigned ϕ values intermediate between ϕ_1 and ϕ_2 on the one hand and ϕ_2 and ϕ_3 on the other. The ϕ values ranged from 0 to 49920, as in the previous example.

Owing to the greater length of the neck in this case, the initial set of ϕ values was not guessed with great accuracy, and more stages of approxi-

Fig. 6.



mation (twenty-four in all) were required before the same degree of convergence in the ϕ values had been obtained, as in the previous case. A slight modification in procedure was adopted, which increased to some extent the rapidity of convergence. In applying the procedure described by equation (4.1), modified (*i. e.* "dashed") values were inserted on the right-hand side of the equation as they were computed, to replace the original (undashed) values. The values of $\frac{\partial \phi}{\partial r}$ and $\frac{\partial \phi}{\partial z}$ were calculated as in the previous example, and the resultant shear stresses were derived from them. The results are illustrated in fig. 6.

For regions remote from the neck, the same general conclusions may be drawn as for the case in which the neck is inclined at 45° to the axis of the

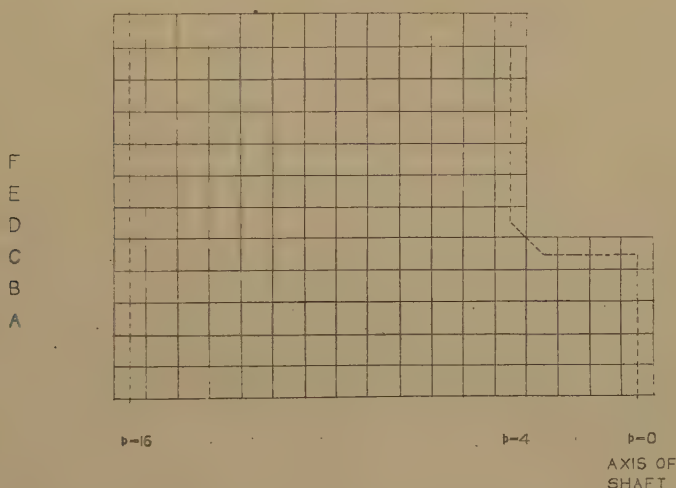
* See also "L. F. R." pp. 314 and 341.

shaft. Within the neck itself, however, the resultant shear stress at the "wide" end has its greatest value near the outer edge, and the same observation may be made for the "narrow" end. The stresses in the neck are much lower than in the preceding example. This is what might have been anticipated in view of the greater length and mean thickness of the neck in this problem, with consequent increase in strength. At the same time it may be noted that the stresses near the points P and Q (fig. 6) are so small that little increase in strength is derived from this region of the neck. In fact, one is led to the conclusion that the whole of the neck might have been inclined to the axis of the shaft at 45° , as in Section V., without serious prejudice to the strength of the shaft in respect of an applied torque.

VII. *Torsion of Shaft with Cylindrical Bore.*

Fig. 7 illustrates half of a meridian section of a uniform cylindrical shaft in which a concentric bore whose diameter is one-quarter that of the

Fig. 7.



original cylinder has been drilled out for a certain length along the axis. The boundaries of the shaft and the bore are shown by the broken lines in the figure. It will be observed that the concave right angles at the end of the bore section have been chamfered at 45° for the reasons given in Section VI. The manner in which the space was sub-divided into square cells is also indicated in this figure.

The arithmetical procedure adopted for this problem was essentially the same as that adopted in the other problems, but the differential

equation (2.1) was transformed, and a different stress function was employed.

If the transformation $\phi = r^{3/2}v$ is effected; then equation (2.1) becomes

$$\frac{\partial^2 v}{\partial r^2} + \frac{\partial^2 v}{\partial z^2} = \frac{15v}{4r^2}. \quad (7.1)$$

In order to solve this equation it may be noted that any plane harmonic function of r and z together with a function v (of r only) satisfying

$$\frac{\partial^2 v}{\partial r^2} = \frac{15v}{4r^2}, \quad (7.2)$$

will give the most general solution.

If the further transformation $r = e^t$ is made and the partial notation is dropped, (7.2) becomes

$$\frac{d^2 v}{dt^2} - \frac{dv}{dt} - \frac{15v}{4} = 0, \quad (7.3)$$

of which the solution is

$$v = Ar^{5/2} + Br^{-3/2}. \quad (7.4)$$

Hence, if f_1 and f_2 are arbitrary functions, then the most general form of v which satisfies (7.1) is

$$v = f_1(r + iz) + f_2(r - iz) + Ar^{5/2} + Br^{-3/2}. \quad (7.5)$$

The corresponding function ϕ is, therefore, in view of the relation $\phi = r^{3/2}v$,

$$\phi = r^{3/2}\{f_1(r + iz) + f_2(r - iz)\} + Ar^4 + B. \quad (7.6)$$

Reference to Section V. shows that the most general solution of the "torsion" equation (2.1) is given by adding any plane harmonic function of r and z , multiplied by $r^{3/2}$, to the solution of the equation at large distances from the neck or varying section. The effect of the plane harmonic function will be marked only in the region in which the section of the shaft is changing. The values of A and B and the forms of f_1 and f_2 will be determined by the form of the varying section and the value of the applied torque.

The finite difference equation corresponding to (7.1) is

$$[p, q] = \frac{4p^2}{16p^2 + 15} [[p+1, q] + [p, q+1] + [p-1, q] + [p, q-1]]. \quad (7.7)$$

Apart from the factor $4p^2/16p^2 + 15$ this is identical with the finite difference form of the plane harmonic equation viz.,

$$[p, q] = \frac{1}{4} [[p+1, q] + [p, q+1] + [p-1, q] + [p, q-1]]. \quad (7.8)$$

Now when $p \geq 10$, $4p^2/16p^2 + 15$ differs from 0.25 by less than 1 per cent. (see table). Thus the values of v at distances from the axis of the shaft greater than one-half its radius (i. e. $p \geq 8$) will differ very little from the values of a function satisfying the harmonic equation and similar

boundary conditions. In performing the successive approximations great simplification is introduced, without any considerable sacrifice of accuracy, by taking $4p^2/16p^2+15$ equal to 0.25, over the whole of the body region, for certainly the first half of the arithmetical work. This simplification, obtained by using equation (7.8) instead of (7.7), enables the arithmetical work to proceed at a much more rapid rate. The sum on the right-hand side of (7.8) was obtained by adding the four values on a calculating machine; it was then divided mentally by four and the new value inserted in the square on which the calculation was centred. As the work progressed and the values of v began to converge, the correct factor ($4p^2/16p^2+15$) was introduced. At first v values for rows near the axis were evaluated by using it, and then it was gradually applied to others at greater distances, until it was in use for every row of squares. The values of this factor are shown in the following table:—

p .	$4p^2/16p^2+15$.	p .	$4p^2/16p^2+15$.
1.....	0.129032	9.....	0.247140
2.....	0.202532	10.....	0.247678
3.....	0.226415	11.....	0.248078
4.....	0.236162	12.....	0.248383
5.....	0.240964	13.....	0.248621
6.....	0.243655	14.....	0.248810
7.....	0.245307	15.....	0.248963
8.....	0.246391		

The initial set of v values was obtained by finding two functions of the type $\phi=Ar^4+B$, one for the solid portion of the shaft and the other for the hollow, and transforming these to v values by using the relation $\phi=r^{3/2}v$. At the first stage the columns C and D (fig. 7) had their v values modified; at the second B, C, D, and E; at the third A, B, C, D, E, and F; and so on in this manner. This problem involved more arithmetical labour than the other two. Although the v values showed considerable variation with r , they varied only slightly with z .

The values of $\frac{\partial \phi}{\partial z}$, which were to be obtained from the small differences between large consecutive ϕ values, were therefore inaccurate unless the values of ϕ (and therefore of v) were determined with greater accuracy, at least than in the previous examples, where ϕ varied more rapidly with z . Accordingly the successive approximations were continued until a degree of convergence of one part in 1000 was obtained in each square. This required thirty stages of approximation. The values of ϕ were then calculated from the final values of v , and the shear stresses calculated

in the same manner as that adopted in the other two problems. The shear stresses are illustrated in figs. 8 and 9. Fig. 8 illustrates their variation in magnitude and direction near the end of the bore, while fig. 9 shows the numerical values of the resultant shear for the remaining squares in those columns which figure in fig. 8.

It is clear from this figure that there is very little departure from linear variation of θz with r even near the end of the hollow portion, and that only comparatively small resultant shears are encountered in this region. The resultant shear shows almost no variation with z for a given r value except near the end of the hollow portion. It is evident, therefore, that at least a quarter of the radius of a uniform shaft may be bored away from part of its length without weakening appreciably its torsional strength. The variations in shear stress near the end of such a bore are of a much smaller order than those which arise near the ends of the necks in the other two examples considered.

VIII. The Convergence of the Sequence Employed.

If u is a function of x and y , and F an arbitrary function of $x, y, u, \frac{\partial u}{\partial x}$ and $\frac{\partial u}{\partial y}$, while J is defined by the equation

$$J = \iint_S F\left(x, y, u, \frac{\partial u}{\partial x}, \frac{\partial u}{\partial y}\right) dx dy, \quad \dots \quad (8.1)$$

where S is some contour in the x, y plane; then the condition that J shall have a minimum value for a certain form of u is

$$\frac{\partial F}{\partial u} - \frac{\partial}{\partial x} \left(\frac{\partial F}{\partial p} \right) - \frac{\partial}{\partial y} \left(\frac{\partial F}{\partial q} \right) = 0, \quad \dots \quad (8.2)$$

where $p = \frac{\partial u}{\partial x}$ and $q = \frac{\partial u}{\partial y}$. This result is known in the Calculus of Variations as Euler's Theorem. On applying this theorem to

$$J = \iint_S \left\{ \left(\frac{\partial u}{\partial x} \right)^2 + \left(\frac{\partial u}{\partial y} \right)^2 \right\} dx dy, \quad \dots \quad (8.3)$$

the minimal condition yields

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0. \quad \dots \quad (8.4)$$

That is to say, the form of u which makes J an absolute minimum is one which satisfies the plane harmonic equation. A similar argument can be applied inversely to the "torsion" equation (2.1), viz.,

$$\frac{\partial^2 \phi}{\partial r^2} - \frac{3}{r} \frac{\partial \phi}{\partial r} + \frac{\partial^2 \phi}{\partial z^2} = 0.$$

Fig. 8.

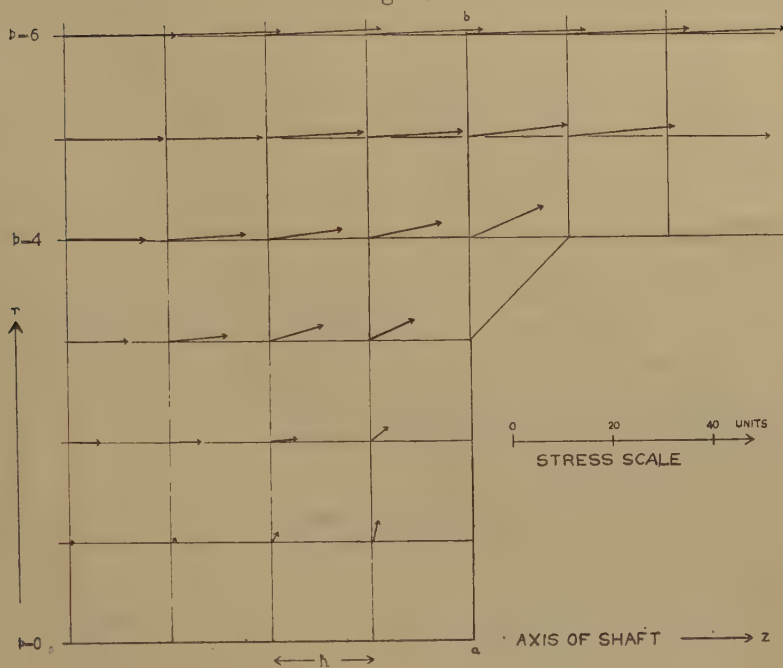


Fig. 9.

560	559	559	557	557	556	556
523	523	523	522	522	522	522
487	487	486	486	487	488	488
453	453	453	453	453	453	453
416	416	417	417	418	419	419
380	380	382	385	384	386	387
343	344	346	347	348	350	352
305	306	308	312	311	313	316
$\frac{p}{7}$ 269	269	270	272	275	277	279

It may be rewritten as

$$\frac{\partial}{\partial r} \left\{ \frac{1}{r^3} \frac{\partial \phi}{\partial r} \right\} + \frac{\partial}{\partial z} \left\{ \frac{1}{r^3} \frac{\partial \phi}{\partial z} \right\} = 0, \quad (8.5)$$

from which it is seen that if

$$J = \iint_s \frac{1}{r^3} \left[\left(\frac{\partial \phi}{\partial r} \right)^2 + \left(\frac{\partial \phi}{\partial z} \right)^2 \right] dr dz, \quad (8.6)$$

then the condition that J should be a minimum is that ϕ should be determined from equation (8.5), *i. e.*, from equation (2.1).

These considerations lead one to deduce by analogy, a "finite difference" form for J , to be denoted by V , and obtained by summation over the body space instead of by integration, which shall stand in the same relation to the finite difference equation (3.5), as does the form of J in (8.6) to the differential equation (2.1). Then the condition that V shall be a minimum may be deduced by direct differentiation, and it will be seen that equation (3.5) results.

Writing as before $\phi(r, s) = [r, s]$ to denote the value of ϕ at the centre of a square whose coordinates are rh, sh , the required form of V is by analogy with (8.6)

$$\begin{aligned} V &= \sum_r \sum_s \frac{1}{r^3 h^3} \left[\left\{ \frac{[r+1, s] - [r-1, s]}{2h} \right\}^2 + \left\{ \frac{[r, s+1] - [r, s-1]}{2h} \right\}^2 \right] \\ &= \frac{1}{4h^5} \sum_r \sum_s \frac{1}{r^3} \{ [r+1, s]^2 + [r-1, s]^2 + [r, s+1]^2 + [r, s-1]^2 \\ &\quad - 2[r+1, s][r-1, s] - 2[r, s+1][r, s-1] \}, \quad (8.7) \end{aligned}$$

where the summation is to be effected for the whole of the body region enclosed by the contour under consideration. Then

$$\begin{aligned} \frac{\partial V}{\partial [v, s]} &= \frac{1}{4h^5} \left[2[r, s] \left\{ \frac{1}{(r-1)^3} + \frac{1}{(r+1)^3} + \frac{2}{r^3} \right\} - 2 \frac{[r-2, s]}{(r-1)^3} \right. \\ &\quad \left. - 2 \frac{[r+2, s]}{(r+1)^3} - 2 \frac{[r, s-2]}{r^3} - 2 \frac{[r, s+2]}{r^3} \right] \\ &= \frac{1}{2h^5 r^3} \left[[r, s] \left\{ 2 + \left(1 - \frac{1}{r} \right)^{-3} + \left(1 + \frac{1}{r} \right)^{-3} \right\} - [r-2, s] \left(1 - \frac{1}{r} \right)^{-3} \right. \\ &\quad \left. - [r+2, s] \left(1 + \frac{1}{r} \right)^{-3} - [r, s-2] - [r, s+2] \right]. \quad (8.8) \end{aligned}$$

If $\frac{1}{r^2}$ and higher powers of $\frac{1}{r}$ are neglected, then

$$\begin{aligned} \frac{\partial V}{\partial [r, s]} &= \frac{1}{2h^5 r^3} \left[4[r, s] - [r-2, s] \left(1 + \frac{3}{r} \right) - [r+2, s] \left(1 - \frac{3}{r} \right) \right. \\ &\quad \left. - [r, s-2] - [r, s+2] \right]. \end{aligned}$$

If, now, the sides of the squares are doubled and the symbols p, q are used to denote the coordinates of the centre of any square, then the above equation may be rewritten as

$$\frac{\partial V}{\partial [p, q]} = \frac{1}{H^5 p^4} [8p[p, q] - [p-1, q](2p+3) - [p+1, q](2p-3) - 2p\{[p, q-1] + [p, q+1]\}]. \quad (8.9)$$

where $H=2h$.

The condition that $\frac{\partial V}{\partial [p, q]}=0$, i.e., the condition that V has a stationary value, is, therefore,

$8p[p, q] = (2p+3)[p-1, q] + (2p-3)[p+1, q] + 2p\{[p, q-1] + [p, q+1]\}$, which is the same relation as (3.5).

It is clear, however, that (3.5) is only an approximation to the minimal condition for V since $\frac{1}{r^2}$ and terms of higher order in $\frac{1}{r}$ have been neglected.

It may be inferred then that (3.5) will yield ϕ values which are an extremely close approximation to those which make V stationary when a large distance from the axis ($\frac{1}{r}$ small) is concerned. If $\frac{1}{r^2}$ is not neglected,

but $\frac{1}{r^3}$ and higher powers of $\frac{1}{r}$ are, then the stationary condition yields other terms in equation (3.5) and the following equation is obtained :

$$8p[p, q] = (2p+3)[p-1, q] + (2p-3)[p+1, q] + 2p\{[p, q-1] + [p, q+1]\} + \frac{6}{2p}\{[p-1, q] + [p+1, q]\}. \quad (8.10)$$

Therefore the error made in $8p[p, q]$ by neglecting $\frac{1}{r^2}$ is of the order of

$$\frac{6}{2p}\{[p-1, q] + [p+1, q]\}.$$

Now $[p-1, q] + [p+1, q] \simeq h^2 \frac{\partial^2 \phi}{\partial r^2} + 2[p, q]^*$.

The proportionate error in $[p, q]$ is therefore given approximately by

$$\frac{3}{8p^2} \left[2 + \frac{h^2 \frac{\partial^2 \phi}{\partial r^2}}{[p, q]} \right].$$

If the second term is small enough to be neglected (and reference to the numbers in fig. 3 shows that

$$\frac{h^2 \frac{\partial^2 \phi}{\partial r^2}}{[p, q]} = \frac{\frac{\partial^2 \phi}{\partial p^2}}{[p, q]}$$

* "L. F. R." p. 309.

varies from about 0.03 to about 1, and that it can be neglected for large values of p) then the error is about $300/4p^2$ per cent. The greatest error which is likely to occur is obtained by putting $p=5$ and the second term equal to 1, whence the percentage error is about 4.5. The stress values near the inner boundaries of the sections in the first two problems are therefore likely to be less accurate than in other regions. On the whole, however, the smaller shear stresses occur nearer the axis, and hence no very great error is produced in the general trend and disposition of the shears.

The value of $\frac{\partial^2 V}{\partial [p, q]^2}$ is found to be $\frac{8}{p^3 H^5} = \frac{8}{H^2 r^3}$, which is essentially positive. Hence the values of ϕ which satisfy equation (8.8) are such as to make V an absolute minimum, *i. e.*, a minimum in respect of variations in any one of the ϕ values. It should be emphasized, however, that the above conclusions are only strictly valid to a degree of approximation which allows $\frac{1}{r^2}$ and higher powers of $\frac{1}{r}$ to be neglected. If such terms are not neglected, then a more complicated finite difference equation results which would yield ultimately a set of body values of greater accuracy (*i. e.*, ones which approximate more closely to the solution of the differential equation (2.1)) than those yielded by the more simple, but less accurate, form of the finite difference minimal condition. It might therefore be expected that when r is small and $\frac{1}{r}$ large, the sequence used will yield inaccurate values of ϕ near the axis of the shaft. It should be noted, however, that, as explained above, the ϕ values for which equation (3.5) is employed are not required for values of p less than 5, and the proportionate errors are not likely to exceed 5 per cent., and will be as high as that only for the smallest stresses which arise. These considerations render the imperfection of the finite difference equation used less significant. If, now, we have a certain set of body values which correspond to a given value of the function V as defined by (8.7), and one of the body values is modified according to equation (4.1), then V is changed to $V + \delta V$, where

$$\begin{aligned} \delta V &\simeq \frac{\partial V}{\partial [p, q]} \delta [p, q] \\ &\simeq \frac{1}{H^5 p^4} [8p\{[p, q] - [p, q]'\}] \\ &\simeq \frac{-8\{\delta [p, q]\}^2}{H^2 r^3} \dots \dots \dots (8.11) \end{aligned}$$

Since only positive values of r arise, δV is essentially negative. Thus repetition of the procedure denoted by equation (4.1) gives a series of decreasing V values. This series will either tend to a limit or the terms

of it will decrease indefinitely. In view of the fact that V is an absolute minimum for a certain set of ϕ values, then the former alternative must be true, and the sequence method will yield ultimately a set of ϕ values which makes V approximately an absolute minimum. Clearly this conclusion is quite independent of the initial set of ϕ values, which may be guessed in the first instance. To shorten the calculations it is however advisable to start with a set of values somewhat near to the true ones, and the methods described in Sections V., VI., and VII. indicate how this may be done.

An exactly similar argument may be applied to the modified equation (7.7), for which the appropriate V function is

$$\sum_p \sum_q \frac{1}{2} [p, q]^2 \left\{ \frac{\sqrt{16p^2 + 15}}{2p} \right\}^2 - [p, q][p, q+1] - [p, q][p+1, q]. \quad (8.12)$$

The condition that V has a stationary value gives

$$\frac{\partial V}{\partial [p, q]} = 0 = [p, q] \cdot \frac{16p^2 + 15}{4p^2} - [p, q+1] - [p, q-1] - [p+1, q] - [p-1, q], \quad (8.13)$$

which is precisely the equation (7.7). One important difference from the above treatment should be noted in this connection. The equation (7.7) follows without any limitation on the value of p or the radial distance. In other words, this finite difference equation (7.7) gives the exact condition that the relevant V function should have a stationary value, whereas equation (3.5) only realizes the corresponding condition approximately. It may be shown also in this case that for ϕ values which satisfy equation (8.13) V is a minimum.

The sequence method used in the solution of the third problem is to be preferred to that used for the first two for two reasons. In the first place equation (8.13) is much better adapted for rapid calculations than equation (3.5); and in the second place it is an exact expression of the minimal condition for V . Equation (8.13) will therefore yield values of ϕ which more nearly satisfy the partial differential equation (2.1) than will those obtained from (3.5).

The same argument may be applied to any partial differential equation of the second order involving two independent variables. In some examples, as for equation (2.1), the resulting finite difference equation may have a complicated form, but some simplification may be obtained by approximating within the finite difference equation itself, as already explained, either at every stage or until some appropriate degree of convergence has been reached. At the same time a suitable transformation of the original differential equation may yield a finite difference equation of greater accuracy in the manner of the transformed equation (7.1).

III. *Experiments in Approximating to Solutions of a Partial Differential Equation.*

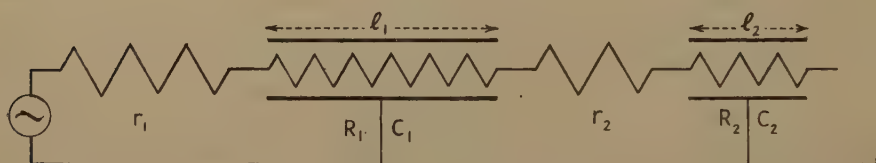
By W. G. BICKLEY, D.Sc.,
Imperial College of Science and Technology *.

[Received August 15, 1940.]

Introduction.

A PROBLEM, emanating from an important section of the electricity supply industry, concerning the effect of varying weather conditions upon the temperature in an average dwelling-house, and the solution of which is expected to assist in obtaining some index of the effect of weather conditions upon the load on the public supply system, was

Fig. 1.



proposed to the author. The problem and the data were given in terms of an analogous electric circuit, shown diagrammatically in fig. 1; the e.m.f. applied is a function of the weather conditions, and the other parameters represent thermal properties of an average room.

An operational solution is not difficult to construct, but its interpretation necessitates the solution of a rather awkward transcendental equation—a matter of considerable labour, and only to be done numerically for each particular set of conditions. It was therefore decided first to try some approximate methods, in the hope that they might afford solutions which were simpler, sufficiently accurate for the purpose in hand, and—with luck—expressible formally in terms of the parameters of the circuit. The possibilities were explored by first applying approximate methods to a pair of simpler circuits, and the results of the experiment seem to be of sufficient interest to warrant publication, on account of the success of some of the attempts, as well as of difficulties experienced with others.

* Communicated by the Author.

or otherwise between the approximations and these exact values is the main criterion we shall apply as to the success of the approximation.

TABLE I.
Values of θ_1 in Problem I, calculated from the Exact and from Approximate Formulæ.

Eqn.	One coefficient.				Exact.	Two coefficients.		
	3.132	3.222	3.243	3.33		4.12	4.22	5.2
τ								
0	0	0	0	0	0	0	0.0385	0
0.2	0.3296	0.4512	0.3812	0.4217	0.2277	0.2338	0.2366	0.2318
0.4	0.5507	0.6988	0.6171	0.6657	0.5255	0.5280	0.5273	0.5318
0.6	0.6988	0.8347	0.7631	0.8067	0.7103	0.7112	0.7090	0.7150
0.8	0.7981	0.9093	0.8534	0.8882	0.8231	0.8233	0.8206	0.8265
1.0	0.8647	0.9502	0.9093	0.9353	0.8920	0.8919	0.8899	0.8943
1.5	0.9502	0.9889	0.9727	0.9836	0.9685	0.9683	0.9673	0.9694
2.0	0.9816	0.9975	0.9918	0.9958	0.9909	0.9907	0.9903	0.9912
2.5	0.9933	0.9994	0.9975	0.9988	0.9973	0.9973	0.9971	0.9974
3.0	0.9975	0.9999	0.9992	0.9997	0.9992	0.9992	0.9992	0.9993

2: *Method of Approximation, with General Equations.*

It is clear that for any $\tau > 0$, θ is a "smooth" function of ξ in the range $0 < \xi < 1$. Such a function can be represented, in this range, with tolerable accuracy, by a polynomial in ξ , of comparatively low degree. The coefficients in this polynomial will be determined by (one or more, linear, ordinary) differential equations, constructed by requiring the assumed solution to satisfy the equation (1.1 *a*) either exactly at a number of points in, or approximately over the whole of, the range (0, 1) of ξ .

We assume that, approximately,

$$\theta = a_0 + a_1\xi + a_2\xi^2 + \dots + a_n\xi^n, \quad \dots \dots \dots (2.1)$$

where the a are functions of τ . Condition (1.21 *a*) gives

$$a_0 = 1, \quad \dots \dots \dots (2.2)$$

while condition (1.23 *a*) gives

$$a_1 + 2a_2 + \dots + na_n = 0. \quad \dots \dots \dots (2.3)$$

From these it follows that

$$\theta = 1 - a_2(2\xi - \xi^2) - a_3(3\xi - \xi^3) - \dots - a_n(n\xi - \xi^n). \quad \dots \dots (2.4)$$

(We have in effect determined a sequence of functions

$$f_r(\xi) = r\xi - \xi^r,$$

each of which satisfies the boundary conditions

$$f(0) = 0, \quad f'(1) = 0.$$

This sequence could have been constructed *ab initio*, and (2.4) then written down.)

We next calculate the "error," i. e., the value of $\partial^2\theta/\partial\xi^2 - \partial\theta/\partial\tau$ for our approximation to θ . It is

$$\begin{aligned} \epsilon &= \partial^2\theta/\partial\xi^2 - \partial\theta/\partial\tau \\ &= 2a_2 + \dot{a}_2(2\xi - \xi^2) + 6a_3\xi + \dot{a}_3(3\xi - \xi^3) + \dots \\ &\quad + n(n-1)a_n\xi^{n-2} + \dot{a}_n(n\xi - \xi^n). \end{aligned} \quad (2.5)$$

3. Approximations with One Disposable Coefficient.

We consider first the case of one coefficient. Possible conditions to impose in order to determine this coefficient as a function of τ , and the sections in which the corresponding results are obtained, are:—

- (3.1) Vanishing of the error at a chosen point of the range. (Collocation.)
- (3.2) Vanishing of the integral of the (possibly weighted) error. (Galerkin method ⁽²⁾.)
- (3.3) Minimizing the square of the error over the range of ξ and τ .

(3.1) Collocation.

The most obvious assumption of this type is to choose a_2 so that the error vanishes at $\xi=1$, the point in which we are most interested. Taking $a_r=0$ ($r>2$), (2.5) then gives

$$2a_2 + \dot{a}_2 = 0, \quad (3.11)$$

whence

$$a_2 = Ae^{-2\tau} \quad (3.121)$$

$$\theta = 1 - (2\xi - \xi^2)Ae^{-2\tau}, \quad (3.122)$$

$$\theta_1 = 1 - Ae^{-2\tau}. \quad (3.123)$$

The constant A is still to be chosen or determined—and condition (1.22 *a*) is still unsatisfied. Nor can this condition be completely satisfied by (3.122); the best we can do is to satisfy it at one point, and as we are particularly interested in θ_1 , we choose as the condition to determine A , that $\theta_1=0$ when $\tau=0$, so that $A=1$ and

$$\theta = 1 - (2\xi - \xi^2)e^{-2\tau}, \quad (3.131)$$

$$\theta_1 = 1 - e^{-2\tau}. \quad (3.132)$$

If we also require that $\theta_1=0$ when $\tau=0$, this leads to

$$\theta=1-(2\xi-\xi^2)e^{-3\tau}, \quad . \quad . \quad . \quad . \quad . \quad (3.221)$$

$$\theta_1=1-e^{-3\tau}. \quad . \quad . \quad . \quad . \quad . \quad (3.222)$$

From the values calculated from (3.222) given in Table I., it is seen that the agreement is generally worse—at least for small and moderate τ —than with (3.132). The explanation of this is not far to seek. As we have already noted, ϵ can vanish at $\xi=0$ only if $a_2=0$, and to compensate for the errors near $\xi=0$ which are thus inevitable, errors of the opposite sign must occur elsewhere in the range (0, 1).

Although the error at $\xi=0$ with non-vanishing a_2 cannot be avoided, its effect can be considerably lessened by “weighting” the error by some function of ξ which vanishes at $\xi=0$. For this, ξ itself, or any positive power of ξ , would serve. Working out the general case with $\xi^n (n>0)$, we have

$$\int_0^1 \{2a_2 + \dot{a}_2(2\xi - \xi^2)\} \xi^n d\xi = 0, \quad . \quad . \quad . \quad . \quad . \quad (3.231)$$

$$\text{i. e.,} \quad \frac{2}{n+1} a_2 + \frac{(n+4)}{(n+2)(n+3)} \dot{a}_2 = 0, \quad . \quad . \quad . \quad . \quad . \quad (3.232)$$

$$\text{leading to} \quad \theta=1-(2\xi-\xi^2)e^{-\alpha\tau} \quad . \quad . \quad . \quad . \quad . \quad (3.241)$$

$$\theta_1=1-e^{-\alpha\tau}. \quad . \quad . \quad . \quad . \quad . \quad (3.242)$$

$$\text{where} \quad \alpha = \frac{2(n+2)(n+3)}{(n+1)(n+4)}.$$

If $n=0$, we reproduce (3.222); if $n=1$, we have

$$\theta_1=1-e^{-2\frac{1}{2}\tau}. \quad . \quad . \quad . \quad . \quad . \quad (3.243)$$

Values obtained from (3.243) are given in Table I.; the index of the exponential is very near to that of the first term of the series in (1.43), but since the coefficient differs considerably, the over-all agreement is hardly so good as with (3.132).

This use of a “weighting” function is essentially equivalent to the method proposed by Galerkin ⁽²⁾.

(3.3) Least Square Error.

It is easy to imagine that a least square error would give the best results obtainable with a single disposable coefficient. We shall see, however, that this is not the case from a practical standpoint, true as it may be theoretically.

We choose a_2 so as to minimize the integral

$$\int_0^\infty d\tau \int_0^1 \{2\bar{a}_2 + a_2(2\xi - \xi^2)\}^2 d\xi.$$

First, carrying out the ξ -integration, we then have to make

$$\delta \int_0^\infty \left(4a_2^2 + \frac{8}{3} a_2 \dot{a}_2 + \frac{8}{15} \dot{a}_2^2 \right) d\tau = 0. \quad (3.31)$$

Cleared of fractions, and simplified, the Eulerian equation is

$$\ddot{a}_2 - 7.5a_2 = 0. \quad (3.32)$$

The positive root of the auxiliary quadratic must be rejected, both on *a priori* physical grounds, and also as not making vanish the integrated portion of (3.31). Making θ_1 vanish at $\tau=0$, we find

$$\theta_1 = 1 - e^{-\tau \sqrt{7.5}} = 1 - e^{-2.7386\tau}. \quad (3.33)$$

From a consideration of the values calculated from (3.33) and given in Table I., it is clear that accuracy for small τ has been sacrificed in order that agreement should be good for large τ . This is a consequence of the infinite range of integration, which could have been foreseen, and is, as we shall see, quite difficult to avoid. An attempt to avoid it would be to apply some weighting function to the square of the error, using a function which tends to zero fairly rapidly with increasing τ . Any prescribed function would be extrinsic and arbitrary, but the a_2 which we seek, and any positive power of it, have the required behaviour. If we then vary

$$\int_0^\infty a_2^n (15a_2^2 + 10a_2 \dot{a}_2 + 2\dot{a}_2^2) d\tau,$$

we obtain the Eulerian equation

$$15(n+2)a_2^{n+1} - (2na_2^{n-1}\dot{a}_2^2 + 4a_2^n \ddot{a}_2) = 0. \quad (3.341)$$

One integration, and the assumption that $a_2 \rightarrow 0$ as $\tau \rightarrow \infty$, give

$$15a_2^{n+2} - 2a_2^n \dot{a}_2^2 = 0. \quad (3.342)$$

This leads again to (3.33) as the relevant solution. We have done no better; no power of a_2 is powerful enough to counteract the effect of the infinite range.

This result would make it appear probable that the least square method is valueless when the range is infinite, unless it is the asymptotic behaviour only which is sought.

(3.4) *Discussion of Results.*

The results so far obtained are far from unsatisfactory. Judged upon the basis of giving a tolerable approximation to θ_1 for small and moderate values of τ , collocation at $\xi=1$ has given the best results so far. This was perhaps to be expected—but expectation is not certainty, nor does it follow that the error in the *solution* of a differential equation is least at the point where the *equation* is most nearly satisfied. The failure of the

“least square” method, applied to an infinite range, again, might perhaps have been foreseen, but it did come rather as a surprise to the author, and it seemed at first incredible that the weighting function tried should be so completely ineffective.

That such results as (3.132) can be obtained with only one disposable coefficient augurs well for the results obtainable with two or more; we proceed with the experiment by obtaining some of these results.

4. Approximations with Two Disposable Coefficients.

Retaining two disposable coefficients— a_2 and a_3 —we have, from (2.5),

$$\epsilon = 2a_2 + \dot{a}_2(2\xi - \xi^2) + 6a_3\xi + \dot{a}_3(3\xi - \xi^3).$$

Discarding “least squares” after the experience of (3.3), we have two possibilities:

- (4.1) Collocation—vanishing of the error at two points in the range (0, 1);
- (4.2) Galerkin method—vanishing of the integral over the range of the weighted error.

We explore these possibilities in turn.

(4.1) Collocation.

To determine two coefficients, we must obtain two equations, and must therefore collocate at two points. The most obvious choice of such points is the extremities of the range, $\xi=0$ and 1. The first of these necessitates $a_2=0$ and the second then gives

$$6a_3 + 2\dot{a}_3 = 0. \quad (4.111)$$

The solution of this, which satisfies the condition (1.22 a) at $\xi=1$ (it cannot also be satisfied at $\xi=0$), makes

$$\theta_1 = 1 - e^{-3\xi}. \quad (4.112)$$

This reproduces (3.222)—but it is to be noted that the value of θ differs from that in (3.221) for all other values of ξ , being here

$$\theta = 1 - \frac{1}{2}(3\xi - \xi^3)e^{-3\xi}. \quad (4.113)$$

If we reject $\xi=0$ as one of our points of collocation—and the vanishing of a_2 may be a valid reason for doing so—there is no other point for which a really good *a priori* reason can be given. The mid-point, $\xi=\frac{1}{2}$, perhaps first suggests itself, and we will next work out this case.

The equations for a_2 and a_3 are

$$2a_2 + \frac{3}{4}\dot{a}_2 + 3a_3 + \frac{11}{8}\dot{a}_3 = 0. \quad (4.121)$$

$$2a_2 + \dot{a}_2 + 6a_3 + 2\dot{a}_3 = 0, \quad (4.122)$$

(4.2) Galerkin Method.

With two (or more) coefficients, it is necessary for their determination that some assumption additional to the vanishing of the integrated (or mean) error be made. The use of a set of weighting functions—essentially the method of Galerkin ⁽²⁾—is evidently a possibility. In our case, with two disposable coefficients, there are three pairs of such functions which seem appropriate. They are :—

- (i.) the first two of the functions used in expanding θ , i. e., 1 and ξ .
- (ii.) the functions associated with the coefficients we seek, in the original expansion, i. e., ξ^2 and ξ^3 ;
- (iii.) the functions associated with these coefficients after the boundary conditions at 0 and 1 have been satisfied, i. e., $(2\xi - \xi^2)$ and $(3\xi - \xi^3)$.

Before we can proceed with any of these, there is a point of procedure which needs consideration and decision, namely, the way in which the condition (1.22 a), when $\tau=0$, is here to be applied. Consistency would suggest that this condition be treated in the same way as the error, although this will almost certainly mean that θ_1 will *not* satisfy the condition. We shall adopt this procedure—and the magnitude of the discrepancy will be one of our experimental results.

For the pair (i.),

$$\int_0^1 \epsilon d\xi = 2a_2 + \frac{2}{3}\dot{a}_2 + 3a_3 + \frac{5}{4}\dot{a}_3 = 0, \quad . \quad . \quad . \quad (4.211)$$

$$\int_0^1 \xi \epsilon d\xi = a_2 + \frac{5}{12}\dot{a}_2 + 2a_3 + \frac{4}{5}\dot{a}_3 = 0, \quad . \quad . \quad . \quad (4.212)$$

along with the conditions at $\tau=0$,

$$\int_0^1 \theta d\xi = 1 - \frac{2}{3}a_2 - \frac{5}{4}a_3 = 0, \quad . \quad . \quad . \quad (4.213)$$

$$\int_0^1 \theta \xi d\xi = \frac{1}{2} - \frac{5}{12}a_2 - \frac{4}{5}a_3 = 0. \quad . \quad . \quad . \quad (4.214)$$

From these we find

$$\theta_1 = 1 - 1.29123e^{-2.48596\tau} + 0.62456e^{-32.1807\tau}. \quad . \quad . \quad (4.215)$$

It is immediately evident that this formula does *not* make θ_1 vanish when $\tau=0$; only collocation (of the boundary condition) at $\xi=1$ will *ensure* this. The *reasons* for the rather large discrepancy are not far to seek; in addition to the fact that ϵ can only vanish at $\xi=0$ if $a_2=0$ (cf. remarks following equation (3.222), the polynomial approximation *cannot* at the same time make $\theta=1$ when $\xi=0$, and $\theta=0$ when $\xi>0$, as the exact solution does, at $\tau=0$. Hence, both in (4.211) and (4.213), large errors near $\xi=0$ must be compensated by large errors near $\xi=1$.

The use of the weighting functions under (ii.) above should, if the above diagnosis is correct, considerably improve the approximation. The result

$$\theta_1 = 1 - 1.24920e^{-2.49839\tau} + 0.28766e^{-19.9562\tau}, \quad \dots \quad (4.22)$$

which is evidently very much better when $\tau=0$. How much it is better for other values of τ may be seen from the values given in Table I. We should also expect that the functions under (iii.) above, which do not reduce so greatly the effect of the inevitable errors near $\xi=0$, would give a result not so good as (4.22), although better than (4.215). The result is

$$\theta_1 = 1 - 1.27940e^{-2.46804\tau} + 0.43325e^{-23.5627\tau}, \quad \dots \quad (4.23)$$

and it agrees with expectation.

(4.3) *Discussion of Results.*

It is clear from the numerical values in Table I. that we get much better agreement with two disposable coefficients than with one—as was to be expected. This is also evident in the formulæ. Each formula in section 4 contains one term which is a more or less successful attempt to approximate to the first term of the series in the exact solution—

$$(4/\pi)e^{-\pi^2/4} = 1.27324e^{-2.46740\tau},$$

and another which is quite a crude attempt to simulate the effect of the rest of the series.

Of the solutions so far obtained, three are outstanding, (3.132), with one coefficient and collocation at $\xi=1$; (4.12), with two coefficients and collation at $\xi=0.5$ and 1; (4.22), with two coefficients, with integrated error and boundary condition, using as weighting functions ξ^2 and ξ^3 .

It seems quite certain that the use of three disposable coefficients would effect considerable further improvement—at the expense, of course, of increased labour.

5. *Three Disposable Coefficients.*

There is one case of three coefficients where the labour is no greater—using a_2 , a_3 , and a_4 , with collocation at 0, 0.5, and 1: for collocation at $\xi=0$ necessitates

$$a_2 = 0, \quad \dots \quad (5.1)$$

so that we are left with a_3 and a_4 only to determine.

The result is

$$\theta_1 = 1 - 1.26306e^{-2.48115\tau} + 0.26306e^{-29.0188\tau}. \quad \dots \quad (5.2)$$

The numerical results given in Table I. may be compared with the exact and with other approximate results.

PART II.—UNIFORM CABLE WITH RESISTANCE AT SENDING END.

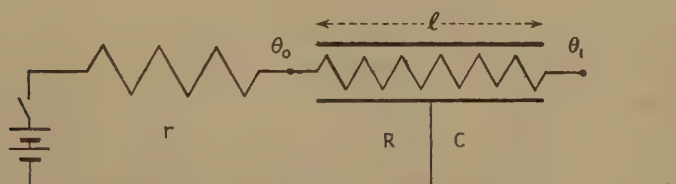
6. Problem II. and its Exact Solution.

The success so far achieved warrants the experiment with a more complicated problem, and we choose as the next subject of experiment the addition of a “lumped” resistance at the “sending end” of the cable—in the thermal analogue, equivalent to including the absorption coefficient for the outer wall, giving the constant whereby the rate of heat absorption per unit area is obtained as a multiple of the temperature difference between the outer air and the external surface of the wall. The circuit is represented diagrammatically in fig. 2, and r is the value of the resistance. Using non-dimensional co-ordinates as before, with the addition of

$$\rho = r/Rl, \quad \dots \dots \dots (6.1)$$

the mathematical specification of the problem is as follows:—

Fig. 2.



To find the solution of the equation

$$\frac{\partial^2 \theta}{\partial \xi^2} = \frac{\partial \theta}{\partial \tau}, \quad \dots \dots \dots (6.2)$$

for which $1 - \theta = -\rho(\partial \theta / \partial \xi) \quad (\xi = 0, \tau > 0). \quad \dots \dots \dots (6.31)$

$\partial \theta / \partial \xi = 0, \quad (\xi = 1, \tau > 0). \quad \dots \dots \dots (6.32)$

$\theta = 0, \quad (0 < \xi \leq 1, \tau = 0). \quad \dots \dots \dots (6.33)$

The exact solution is known. If λ_n denote the n th positive root of the equation

$$\cot \phi = \rho \phi, \quad \dots \dots \dots (6.4)$$

then $\theta = 1 - 2 \sum_{n=1}^{\infty} \frac{e^{-\lambda_n^2 \tau}}{\lambda_n} \frac{\cos \lambda_n (1 - \xi)}{(1 + \rho + \rho \lambda_n^2) \sin \lambda_n}, \quad \dots \dots \dots (6.51)$

$$\theta_0 = 1 - 2 \sum_{n=1}^{\infty} \frac{\rho e^{-\lambda_n^2 \tau}}{(1 + \rho + \rho \lambda_n^2)}, \quad \dots \dots \dots (6.52)$$

$$\theta_1 = 1 - 2 \sum_{n=1}^{\infty} \frac{\rho e^{-\lambda_n^2 \tau}}{(1 + \rho + \rho \lambda_n^2) \cos \lambda_n} \quad \dots \dots \dots (6.53)$$

In the numerical examples of this part we shall use the value $\rho=1$. Early values of the λ_n , for this value of ρ , are

$\lambda_1=0.86038 \dots, \quad \lambda_2=3.42580 \dots, \quad \lambda_3=6.43733 \dots;$
 $\lambda_4=9.52934 \dots,$

while values of λ_{n+1} for large n approximate to $n\pi$, or, more accurately,

$$\lambda_{n+1}=n\pi+\frac{1}{n\pi}-\frac{4}{3(n\pi)^3}-\frac{47}{15(n\pi)^5}\dots$$

Values of θ_0 and θ_1 for a range of values of τ are given in Table II.

TABLE II.

Values of θ_0 and θ_1 in Problem II, calculated from the Exact and from Approximate Formulæ.

	Eqn.	One coefficient.		Exact.	Two coefficients.		
		8.181	8.241		9.121	9.221	9.231
$\theta_0 \dots\dots$	τ						
	0.0	0.0	0.2500	0.0	0.0	0.0	0.1111
	0.2	0.1813	0.3545	0.3566	0.3255	0.3001	0.3553
	0.4	0.3297	0.4444	0.4559	0.4326	0.4113	0.4552
	0.6	0.4512	0.5218	0.5317	0.5158	0.4940	0.5312
	0.8	0.5507	0.5884	0.5963	0.5863	0.5635	0.5958
	1.0	0.6321	0.6457	0.6518	0.6472	0.6236	0.6515
	1.5	0.7769	0.7565	0.7596	0.7619	0.7401	0.7593
	2.0	0.8647	0.8321	0.8339	0.8395	0.8205	0.8338
	2.5	0.9179	0.8850	0.8853	0.8918	0.8761	0.8853
	3.0	0.9502	0.9209	0.9208	0.9271	0.9144	0.9208
$\theta_1 \dots\dots$	4.0	0.9817	0.9627	0.9622	0.9669	0.9592	0.9622
	5.0	0.9933	0.9824	0.9819	0.9849	0.9806	0.9820
	Eqn.	8.142	8.242	..	9.122	9.222	9.232
	τ						
	0.0	0.0	-0.125	0.0	0.0	0.0	0.0370
	0.2	0.1248	+0.0317	0.0494	-0.0110	-0.0247	0.0478
	0.4	0.2358	0.1668	0.1690	+0.1300	+0.1017	0.1670
	0.6	0.3297	0.2827	0.2822	0.2566	0.2240	0.2815
	0.8	0.4160	0.3826	0.3810	0.3651	0.3308	0.3803
	1.0	0.4866	0.4686	0.4661	0.4578	0.4229	0.4656
	1.5	0.6321	0.6348	0.6313	0.6345	0.6015	0.6310
	2.0	0.7364	0.7490	0.7453	0.7537	0.7248	0.7452
	2.5	0.8111	0.8275	0.8241	0.8332	0.8100	0.8241
	3.0	0.8647	0.8814	0.8785	0.8881	0.8688	0.8785
	4.0	0.9305	0.9440	0.9420	0.9491	0.9375	0.9421
	5.0	0.9644	0.9735	0.9723	0.9769	0.9702	0.9724

7. Approximate Solution ; General Equations.

As before, we assume a polynomial as an approximation to θ as a function of ξ , the coefficients being functions of τ ,

$$\theta = a_0 + a_1\xi + a_2\xi^2 + a_3\xi^3 + \dots + a_n\xi^n. \quad (7.1)$$

The two boundary conditions (6.31) and (6.32) lead respectively to

$$1 - a_0 = -\rho a_1, \quad (7.21)$$

$$a_1 + 2a_2 + \dots + na_n = 0. \quad (7.22)$$

Hence,

$$\theta = 1 - a_2(2\rho + 2\xi - \xi^2) - a_3(3\rho + 3\xi - \xi^3) \dots - a_n(n\rho + n\xi - \xi^n), \quad (7.31)$$

$$\theta_0 = 1 - 2\rho a_1 - 3\rho a_2 - \dots - n\rho a_n, \quad (7.32)$$

$$\theta_1 = 1 - (2\rho + 1)a_2 - (3\rho + 2)a_3 - \dots - (n\rho + n - 1)a_n. \quad (7.33)$$

With this value of θ the error in the differential equation is

$$\epsilon = \{2a_2 + (2\rho + 2\xi + \xi^2)\dot{a}_2\} + \{6\xi a_3 + (3\rho + 3\xi - \xi^3)\dot{a}_3\} \dots \\ + \{n(n-1)\xi^{n-2}a_n + (n\rho + n\xi - \xi^n)\dot{a}_n\}. \quad (7.4)$$

Approximation will be made by methods similar to those used in Part I ; the criterion of success will be the agreement of the approximate with the accurate values of θ_0 and θ_1 ; in the practical problem, interest in θ_1 would be much greater than that in θ_0 .

8. Approximations with One Disposable Coefficient.

(8.1) Collocation.

As a first experiment we collocate at $\xi=1$. This gives

$$2a_2 + (2\rho + 1)\dot{a}_2 = 0, \quad (8.11)$$

whence $a_2 = Ae^{-\alpha\tau}, \quad (\alpha = 2/(2\rho + 1)), \quad (8.12)$

$$\theta_0 = 1 - 2\rho Ae^{-\alpha\tau}, \quad (8.131)$$

$$\theta_1 = 1 - (2\rho + 1)Ae^{-\alpha\tau}. \quad (8.132)$$

We have still to apply (6.33)—and it is evident that the same value of A will not allow it to be satisfied at both ends of the cable ; this difficulty is going to occur in all the cases in which we have only one disposable coefficient. If we choose to satisfy it at $\xi=1$, we have $A=1/(2\rho+1)$, so that

$$\theta_0 = 1 - 2\rho e^{-\alpha\tau}/(2\rho + 1), \quad (8.141)$$

$$\theta_1 = 1 - e^{-\alpha\tau}. \quad (8.142)$$

Values of θ_1 , for $\rho=1$, are given in Table II.

If, on the other hand, we prefer to collocate at $\xi=0$, we obtain

$$2a_2 + 2\rho\dot{a}_2 = 0, \quad (8.15)$$

(9.2) *Galerkin Method.*

We will next obtain the result of using the Galerkin method with weighting functions 1 and ξ . It may at first sight seem that the second of these privileges the region in the neighbourhood of $\xi=1$. If we reflect, however, that the weighting function $(1-\xi)$ replacing ξ would lead to the same results, this hasty inference is seen to be fallacious.

The equations for the coefficients are

$$2a_2 + \left(2\rho + \frac{2}{3}\right)\dot{a}_2 + 3a_3 + \left(3\rho + \frac{5}{4}\right)\dot{a}_3 = 0, \quad \dots \quad (9.211)$$

$$a_2 + \left(\rho + \frac{5}{12}\right)\dot{a}_2 + 2a_3 + \left(\frac{3}{2}\rho + \frac{4}{5}\right)\dot{a}_3 = 0. \quad \dots \quad (9.212)$$

Slightly different results emerge according as we treat the initial conditions by collocation, or by the Galerkin method. If the former, collocating at 0 and 1, we have

$$\begin{aligned} \theta_0 &= (19 - 15e^{-\alpha\tau} - 4e^{-\beta\tau})/19 \\ &= 1 - 0.78947e^{-\alpha\tau} - 0.21053e^{-\beta\tau}, \quad \dots \quad (9.221) \end{aligned}$$

$$\begin{aligned} \theta_1 &= (19 - 23e^{-\alpha\tau} + 4e^{-\beta\tau})/19 \\ &= 1 - 1.21053e^{-\alpha\tau} + 0.21053e^{-\beta\tau}, \quad \dots \quad (9.222) \end{aligned}$$

where

$$\alpha = 20/27 = 0.74074 \dots, \quad \beta = 12.$$

If, alternatively, the initial condition is treated by the Galerkin method with the same weighting functions, we find

$$\theta_0 = 1 - 0.73099e^{-\alpha\tau} - 0.15789e^{-\beta\tau}, \quad \dots \quad (9.231)$$

$$\theta_1 = 1 - 1.12086e^{-\alpha\tau} + 0.15789e^{-\beta\tau}, \quad \dots \quad (9.232)$$

where α and β have the same values as above.

These latter suffer from the disadvantage of our criterion, since they do not satisfy the initial conditions at the ends of the ξ -range; for greater values of τ , however, both sets give good agreement with the accurate solution.

10. *Summary and Conclusions.*

The paper has described a number of experimental attempts to approximate to the solution of the differential equation

$$\frac{\partial^2 \theta}{\partial \xi^2} = \frac{\partial \theta}{\partial \tau}$$

for the ranges $0 < \xi < 1$, $0 < \tau < \infty$, with certain boundary and initial conditions. By assuming a polynomial in ξ , the coefficients of which are functions of τ , the problem is reduced to solving ordinary differential equations for these coefficients. These differential equations may be

constructed by collocation, or by the Galerkin method of (weighted) integrals. The initial conditions are similarly treated. Quite good agreement is obtained with only two disposable coefficients; if interest is largely centred in one point, collocation at that point produces results better than the Galerkin method. The method of least squares applied to the error is apparently not well suited for use with an infinite range, unless it is the asymptotic behaviour only which is sought.

References.

- (1) See, for instance,
Carson, 'Electrical Circuit Theory and Operational Calculus,' pp. 15-16.
McLachlan, 'Complex Variable and Operational Calculus,' Appendix 8,
pp. 321-325, and p. 168 *et seq.*
Warren, 'Mathematics applied to Electrical Engineering,' pp. 313-315.
- (2) Duncan, 'Galerkin's Method in Mechanics and Differential Equations,'
Aeronautical Research Committee, R. and M. 1798 (1937); and
'The Principles of the Galerkin Method,' *ibid.* R. and M. 1848 (1938).

IV. *Friction and Adhesion.*

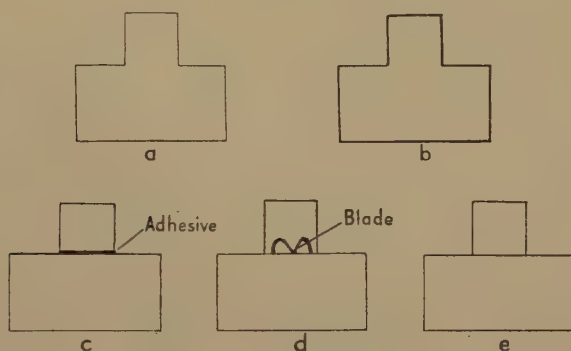
By J. J. BIKERMAN*.

[Received February 10, 1941.]

1. "*Resistance to Sliding*" and "*Friction.*"

THE aim of this paper is to show that friction is not due to adhesion. Many experiments were performed to support the opposite view, *i. e.*, to prove the identity between adhesional and frictional forces. In order to discuss them it is unfortunately necessary to define "friction" more precisely than is usual †.

Fig. 1.



A body shaped like *a* in fig. 1 may possess one of many physical structures. Four of them which are of importance for the theory of friction are illustrated by sketches *b* to *e*. In *b* the cube and the plate form a solid unit without joints. In *c* the cube is glued to the plate. In *d* the cube is provided with a blade invisible from without. In *e* the cube lies freely on the plate.

* Communicated by the Author.

† Holm and Kirschstein ⁽¹⁾ state: "we use the term static friction for the limiting value of the tangential force as long as it is unambiguously determined by the acting normal pressure. On the other hand, sticking is present when the tangential force depends also on the previous pressures if they were higher."

The friction is usually defined as the resistance to sliding. If the cube is forced to slide past the surface of the plate, in case *b* the ultimate shearing strength of the material would be measured, in case *c* the strength of the adhesive joint, in case *d* the scratch hardness of the plate, and in case *e* the real friction.

If it is impossible to decide by simpler means to which case a given system belongs, the clue may be found by observing its behaviour when tangential forces are applied. In case *b* the resistance to sliding is not reproducible; after the cube has been moved away and returned to its original position the force necessary to induce a second sliding is quite different from that required at the first attempt. In case *c* it depends on the adhesive whether the results are, or are not, reproducible. In case *d* the resistance is not reproducible if the blade exactly resumes its original position, but it will have its original magnitude if the new position differs however slightly from the original one. When in cases *c* and *d* the resistance to sliding appears reproducible it still can be distinguished from case *e*, since it does not comply with Coulomb's (or Amonton's) law of friction. This law affirms first that the resistance to sliding is proportional to the normal force. The force required to break an adhesive layer or to produce a scratch may, under special circumstances and within some limits, be proportional to the load on the cube or the scratching needle, but the validity range will be much less extensive than that of Coulomb's law, and the breakdown of the proportionality between tangential and normal forces will indicate the presence of case *c* or *d* as distinct from *e*. The second half of Coulomb's law requires the resistance to sliding to be independent of the area of contact. This rule, of course, never applies to adhesion or to scratching, and allows an even easier decision as to whether a given system belongs to one of the types *c* and *d* or to *e*. In case *e* the resistance to sliding is reproducible, and both components of Coulomb's law are valid.

It is, of course, a matter of personal taste how friction is defined. It is possible to define it so as to include two, three, or all the four cases considered above. But then the "friction" has no general cause and cannot be accounted for by any single mechanism. Then there exists also no "law of friction," as the relationship between tangential force, load, and contact area is different in cases *b* to *e*. It is not correct to examine, say, the mechanism of case *d* and then deduce from this mechanism the law of Coulomb, which does not apply to this case.

2. *Discussion of some Experiments on Resistance to Sliding.*

Several publications on sliding friction deal in reality with other forms of the resistance to sliding; some, unfortunately, disregard this fundamental difference.

Shearing strength (case *b*) presumably is determined when relative sliding of two bodies roughened by repeated heating and cooling is observed *in vacuo*. Under these conditions the reproducibility of the results was poor ^(1, 2). The angle of sliding was very high. In some cases there was no sliding even at the tilt of 90° ; when the system was slightly tapped the slider fell and sometimes stuck at some other point of the supporting surface: the effect was similar to, and presumably had the same mechanism as, say, the fall of a bough broken by wind in a forest; it was therefore certainly different from the real friction. The "friction" of an anchor entangled in the sea bottom also mostly belongs to this type. Even when there was a regular sliding ⁽²⁾ it violated the law of Coulomb. For instance, the apparent coefficient of friction between aluminium was reduced from about 3 to about 1.3 when the load was raised from 30 to 180 g.

Adhesion (case *c*) is operative, for instance, in wet air when the capillary attraction produced by water condensed between slider and support increases the normal reaction. An experimental and theoretical investigation of this attraction was published by Macaulay ⁽³⁾. As the attraction, although it increased with increasing load, was not proportional to it, Coulomb's law was again invalid. The apparent coefficient of friction between wet glass surfaces decreased from 0.56 at 2 g. to 0.35 at 15 g. load; dried (and slightly greased) glass gave 0.20–0.21 at 2–7 g. and 0.27–0.28 at 9–21 g. load. The results were reproducible since here the "adhesive" assumed its final strength immediately after the contact was made.

Resistance to scratching (case *d*) has been repeatedly measured and mistaken for friction. Rudolph Franz ⁽⁴⁾ was apparently the first to measure the force *F* required to move a plate under a scratching needle. He used this force to estimate the scratch hardness of gypsum in different crystallographic directions. The higher the hardness the smaller the force required since the width and depth of the scratch are smaller as well. Since higher load produces deeper scratches, *F* increases with load *W*, but the quantitative relation between *F* and *W* (as far as could be ascertained) has never been determined systematically. A discussion of scratch-hardness measurements ("sclerometry") may be found in O'Neill's monograph ⁽⁵⁾.

Hardy and Hardy ⁽⁶⁾ applied Franz's method to glass. The scratches produced were easily visible and resembled scratches made, for instance, by cutting diamond ⁽⁷⁾. Since scratch hardness is almost independent of the degree of roughness of the surface, Hardy and Hardy found the force *F* to be identical for optically polished and for ordinary plate-glass. The ratio *F*/*W* was independent of *W* when *W* varied as 1 : 3, but the second part of Coulomb's law was not checked, *i. e.*, the area of contact

was not varied *. In contrast with the real coefficient of friction the ratio F/W was, in experiments of Hardy and Doubleday ⁽⁸⁾, strongly and irreversibly affected by a small rise of temperature.

Tomlinson ⁽¹⁰⁾ used Franz's method for glass and several metals, and confirmed the close connection between the ratio F/W and hardness. He expressed his results by the equation $F/W = \text{const.} (\vartheta_A + \vartheta_B)^{\frac{1}{2}}$, ϑ_A and ϑ_B being Auerbach's indentation moduli of the materials A and B in contact. No check was made of whether any of the requirements of Coulomb's law was met, *i. e.*, neither the load was varied nor a plane slider substituted for the needle.

No attempt to test the applicability of Coulomb's law was made by Bowden and Leben ⁽¹¹⁾ either, who likewise used a modification of Franz's method. A needle used produced on steel a scratch 0.002 cm. wide. The paper contains photographs of scratches and admission that some metals "plough out a groove."

Ernst and Merchant ⁽¹²⁾ determined the resistance to sliding in a range of pressures, including the transition from friction to scratching. Unfortunately, data are given in a graph for two examples only. At small loads (8 g. for the pair iron/iron and 12 g. for antimony/antimony) Coulomb's law was apparently still valid; the coefficient of friction was 0.4 for iron/iron and 0.28 for antimony/antimony. When the load W was increased the force F increased more rapidly than W , and the ratio F/W reached 1.1 for iron (at 25 g. load) and 0.5 for antimony (at 40 g. load). This ratio appeared to be independent of the load at higher loads and in so far was similar to the coefficient of friction; but the latter is independent of the area of contact, and this point has not been tested.

Scratch hardness depends but little on the conditions of the surface and is related to bulk properties of the material. So is, of course, the ratio F/W determined by Franz's method. As Tomlinson, Bowden and Leben, and Ernst and Merchant believed F/W to be the coefficient of friction, they concluded that "the friction is certainly not a surface effect only," and is a function of the bulk properties of the bodies in contact.

According to Tomlinson these properties were modulus of rigidity and modulus of compressibility (or modulus of elasticity and Poisson's ratio). A correlation between compressibility and rigidity on one hand and hardness on the other hand had to be expected.

Bowden and Leben had the impression that the ratio F/W was linked with the melting point. A relation between melting point and hardness

* For deliberately lubricated surfaces Hardy and Bircumshaw ⁽⁹⁾ found a difference between the values of F/W observed with curved and plane steel sliders.

was suggested before ⁽¹³⁾. The ratio F/W appeared to be also a function of the indentation hardness; an affinity between indentation hardness and scratch hardness is well known.

It is hardly necessary to point out that the real coefficient of friction, *i. e.*, the factor μ in Coulomb's equation $f = \mu W$ (f is frictional force and W load, μ is independent of W and area of contact) is no function of hardness, rigidity, melting point, etc. All these properties are identical for quartz sand and for polished quartz, but μ may be 0.7 for sand ⁽¹⁴⁾ and 0.3 for polished quartz. All these quantities are different for glass and brass, but the value of μ may be identical for both these materials ⁽¹⁵⁾.

3. Discussion of Adhesion Theories of Friction.

Coulomb's theory identifies the frictional force with that necessary to lift the slider over the asperities of the supporting surface; no molecular interactions are taken into account. Hardy and Doubleday ⁽¹⁶⁾ rejected this theory and insisted on a "molecular" mechanism of friction: "The surface of a solid is the locus of a powerful field of attraction which is the cause of capillary phenomena, and this attraction, though it may be modified by the film of contamination on all naturally occurring solid faces, is not destroyed by it, otherwise, for example, a drop of water would not cling to a glass rod." An evident reply to this reasoning is: since there is an attraction between glass and water, the coefficient of slipping is infinitely large, *i. e.*, there is no slipping of water at the walls of a glass tube in which it flows. The absence of slipping persists however small may be the pressure of water or its amount ("load"). These are typical features of an adhesion. But the coefficient of friction is not infinitely large, and the frictional force is proportional to the load. Why this contrast in behaviour, if the mechanisms of friction are identical for solid/solid and for solid/liquid interfaces? The correct explanation is, of course, that the mechanisms are completely different: no slipping occurs since water displaces air from the glass surface, and sliding takes place since solids cannot displace air from other solid surfaces; no mutual wetting of solids is possible.

Hardy and Hardy ⁽⁶⁾ and Hardy and Doubleday ⁽¹⁶⁾ suggested that friction was "due to molecular attraction operating across an interface, in short, to cohesion." This hypothesis implies the identity of case *e* with cases *b* * and *c*. It is inconsistent with the difference observed experimentally between these cases when *tangential* forces operate (see section 1). It also contradicts the behaviour of *normal* forces. If, in case *b*, the cube has to be separated from the plate, the tensile strength of the material has to be overcome. In case *c* as well the lifting of the cube would require a force which is higher than its weight, namely

* See also R. Schnurmann ⁽¹⁸⁾.

that necessary to overcome the adhesion. When Coulomb's law is valid, the minimum force sufficient to lift the slider is its weight; therefore the mechanism suggested by Hardy does not operate. Hardy and Hardy already raised the question as to why the adhesion postulated by them could not be detected by measuring the normal forces. Two tentative explanations were given by Hardy and Hardy. The first is rather indefinite, and when an attempt is made to formulate it more precisely it seems to conflict with the present knowledge. The other explanation suggests that friction is due to an irreversible rotation of surface molecules produced by relative motion of two surfaces and, hence, is different from cohesion, which can be overcome without changing the orientation of molecules. Unless a very special model is imagined, this assumption is equivalent to saying that the shearing strength of all bodies is finite and their tensile strength is infinitely small. In the later paper⁽¹⁷⁾ summarising Hardy's friction work these speculations have not been repeated; on the contrary, it was admitted that the hypothesis implying "that friction . . . was some direct function of cohesion . . . certainly is not proven, for what is known of the tangential and normal reactions presents a striking picture of contrasts." Apparently the literature contains no other attempt to explain the absence of a detectable adhesion between slider and support.

In a note by Macaulay⁽¹⁹⁾, Hardy's cohesion was intensified to fusion across the surface layers. Later⁽¹¹⁾ it was termed welding. It is apparently not sufficiently realised that not only the form of the law of friction but its numerical constant as well may be identical for the pairs platinum/platinum, wood/wood, and platinum/wood. Unless one is prepared to speak of fusion (or welding) between platinum and wood, fusion cannot be the cause of friction. Besides, any welded joint-like cohesion would hinder the normal separation of slider and support; and this hindrance—when Coulomb's law is valid—does not exist.

Hardy himself recognized that the cohesion theory of friction is also unable to account for Coulomb's law; adhesion force would be proportional to the area of contact whilst friction is independent of the latter. This difficulty seemed to be overcome when Tomlinson⁽¹⁰⁾ pointed out that the adhesion was proportional to the actual area A of contact as distinct from the geometrical area. As the actual area is so far unknown, it is possible to assume its proportionality with the load: $A = \nu W$, ν being a constant factor. In conjunction with Coulomb's empirical law, this hypothesis would give for the frictional force f the equation

$$f = \frac{\mu}{\nu} A, \quad (1)$$

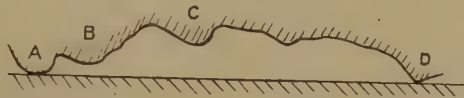
in agreement with the adhesion theory of friction.

The tentative justifications of Tomlinson's hypothesis advanced by Tomlinson and by Bowden and Tabor ⁽²⁰⁾ are conflicting and fail to convince; it would be too long to discuss them here*.

Equation (1) was disproved by simple experiments ⁽¹⁵⁾. Suppose a frictional force f_1 has to be overcome when a slider weighing W_1 g. is moved. If the slider is loaded with an additional weight W_2 a frictional force f_2 is active. If $f_1 = \mu_1 W_1$ and $f_2 = \mu_2 (W_2 + W_1)$ then, according to Coulomb's law, $\mu_1 = \mu_2$. Now the additional load is removed and the frictional force f_3 for slider alone is determined. If there is no adhesion, and the deformation of surface is elastic, then $f_3 = f_1$. Experiments showed that f_3 was equal to f_1 as long as μ_1 was equal to μ_2 , *i. e.*, for surfaces fulfilling Coulomb's law friction was not due to adhesion.

As the validity of this reasoning was questioned ⁽¹²⁾ it is worth while to repeat it in another way. Taking the point of view of the adhesion theory, when slider W_1 is placed on the support, the actual area of contact is νW_1 . When W_2 is added on top of W_1 the actual area of contact is $\nu (W_1 + W_2)$. When W_2 is removed again the area will depend on whether

Fig. 2.



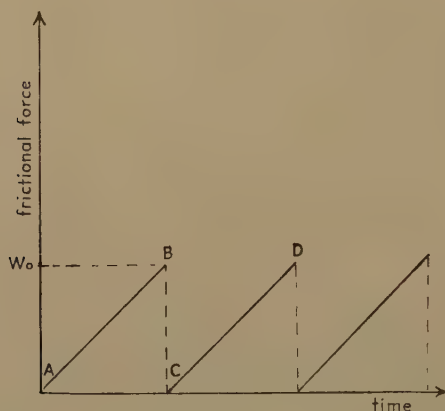
the deformation was chiefly plastic or purely elastic. For a plastic deformation this area evidently will be equal to $\nu (W_1 + W_2)$; this follows from the definition of plastic deformations. Since adhesion is proportional to area the force f_3 required to overcome the adhesion would be equal to f_2 . In reality it is equal to f_1 . For an elastic deformation the area of contact, after removal of the additional load, would be less than $\nu (W_1 + W_2)$ but larger than νW_1 . It would be the nearer to $\nu (W_1 + W_2)$ the stronger the adhesion; in general, it also would be the nearer to $\nu (W_1 + W_2)$ the higher the coefficient of friction. Fig. 2 may help to verify it. When no additional weight is present the bodies in contact touch over the areas A and D. When W_2 is added the arc ABCD may be bent so as to bring the lower surface into contact with the areas B and C as well. When the weight W_2 is removed the arc tends to return to its original position, but is prevented to do so by the welded joints B and C. The elasticity may be strong enough to break some of the joints (*e. g.* C) but it cannot

* According to Bowden and Tabor's own calculation the real area of contact increases in the ratio 17000 : 13, *i. e.* 1300 : 1, when the applied load increases in the ratio 100 : 1. There is evidently no proportionality between A and W.

break all of them unless there is no welding (or adhesion) at all. This may be best shown by assuming the slider being attached to a balance beam. If, after W_2 has been removed, weights are placed on the other pan of the balance, the load on the boundary between slider and support decreases, and, according to the hypothesis, the actual area of contact decreases in the same ratio. If the weights on the other pan reach the value of W_1 the load and the actual area of contact would be equal to nothing. In other words, a weight W_1^* would be sufficient to lift the slider weighing W_1 g.; hence there was no adhesion between slider and support. If adhesion is present the actual area of contact cannot be reduced to nothing by decreasing the load.

Some recent experiments by Bowden and Leben⁽¹¹⁾ were thought to be best accounted for by the welding theory of friction. They may

Fig. 3.



be recorded as follows:—Let us consider the usual device for measuring friction, *i. e.*, a slider pulled by a cord passing over a pulley to a weight W . When W is gradually increased the frictional force f remains equal to W as long as there is no sliding. Plotting f on the ordinate and W on the abscissa a straight line is obtained. If the weight W is increased at a constant speed, *i. e.*, if $W = \text{const. } T$, there will be a linear relation between f and the time T as well (see AB in fig. 3). When W reaches the value W_0 of the limiting frictional force, a slip occurs. Then the slider is put back and the experiment repeated. If the time spent on putting the slider back and removing weights is short as compared with the time used to raise the weight from nothing to W_0 , the new line CD will start

* Neglecting the friction of the balance beam, etc.

almost exactly underneath point B. The variation of the frictional force with time will consist of a series of parallel lines inclined to the time axis and connected by lines almost parallel to the f -axis. Bowden and Leben termed this pattern "stick-slips".* In their experiments the pulling force was increased continuously as distinct from discrete increases by adding small weights; they used a spring which pulled the slider, and the tension of the spring was continuously changed †. It is obvious that no conclusions regarding the mechanism of friction can be drawn from this new representation of frictional experiments, however elegant it may be.

4. Coulomb's Theory.

It was shown above that friction, as long as it fulfills Coulomb's law, cannot be due to adhesion of any kind. It is worth while to emphasize that Coulomb's theory attributing friction to the roughness of surfaces in contact gives the correct order of magnitude of the coefficient of friction. This coefficient μ is a pure number, and *a priori*, there is no reason to suppose that it would be somewhere between $1 \cdot 10^{-1}$ and $5 \cdot 10^{-1}$; the values, say, $1 \cdot 10^{-3}$ to $5 \cdot 10^{-3}$ or $1 \cdot 10$ to $5 \cdot 10$ would be as probable. On Coulomb's view, μ is approximately equal to the tangents of the effective angle between asperities and the basic plane. If μ were, say, $1 \cdot 10^{-3}$ to $5 \cdot 10^{-3}$, this angle would be $4'$ to $18'$; if μ were $1 \cdot 10$ to $5 \cdot 10$ this angle would be 84° to 87° . These values would be incompatible with the known roughness of solid surfaces⁽²²⁾; indeed, they would be sufficient to refute Coulomb's theory. But the observed values require the effective slope of asperities to be 6° – 26° . Their meaning is that it is very difficult to prepare a surface on which no facet is inclined relatively to the basic plane at less than, say, 6° ; and, on the other hand, that surfaces showing irregularities steeper than 26° are avoided in frictional experiments. Bodies having no smooth surfaces, like "earth" of friction tables, may have values of over 1.

Summary.

Recent literature on sliding friction is critically reviewed. It is concluded that:

(1) Friction cannot be due to welding as the law of friction and its numerical constant may be identical for the pairs Pt/Pt, Pt/wood, and wood/wood.

* Some peculiarities of the "stick-slip" curves were due to the fact that mostly surface hardness as distinct from real friction was measured.

† A more detailed interpretation appears to be given elsewhere⁽²¹⁾. There the mechanical conditions are mentioned which favour the rhythmic movement of the slider; they are a low mass of the slider, a low rate of increase of the spring tension, a high static friction, etc.

(2) Friction cannot be due to adhesion in general as no adhesion can be detected by measuring the normal force.

(3) Coulomb's law cannot be reconciled with the adhesion hypothesis by assuming the real area of contact to be proportional to the load; this assumption is refuted by experiments using decreasing loads.

(4) In experiments which seemed to prove that friction was not due to the surface roughness not the real friction was measured; mostly it was scratch hardness.

(5) "Stick-and-slip" patterns have no bearing on the mechanism of friction.

(6) The absolute values of the coefficient of friction are accounted for by Coulomb's theory attributing friction to surface asperities.

References.

- (1) R. Holm and B. Kirschstein, *Wiss. Veroeff. Siemens-Werken*, xv. p. 122 (1936).
- (2) P. E. Shaw and E. W. L. Leavey, *Phil. Mag.* (7) x. p. 809 (1930).
- (3) J. M. Macaulay, 'Engineering,' Nov. 1926.
- (4) R. Franz, *Pogg. Ann. d. Physik*. xx. p. 37 (1850).
- (5) H. O'Neill, 'The Hardness of Metals and its Measurement,' London, pp. 135-150 (1934).
- (6) W. B. Hardy and J. K. Hardy, *Phil. Mag.* (6) xxxviii. p. 32 (1919).
- (7) A. J. Holland and W. E. S. Turner, *J. Soc. Glass Techn. (Trans.)*, xxi. p. 383 (1937).
- (8) W. B. Hardy and I. Doubleday, *Proc. Roy. Soc. A*, ci. p. 487 (1922).
- (9) W. B. Hardy and I. Bircumshaw, *Proc. Roy. Soc. A*, cviii. p. 1 (1925).
- (10) G. A. Tomlinson, *Phil. Mag.* (7) vii. p. 905 (1929).
- (11) F. P. Bowden and L. Leben, *Proc. Roy. Soc. A*, clxix. p. 371 (1939).
- (12) H. Ernst and M. E. Merchant, *Proceedings of the Conference on Friction*, Massachusetts Inst. of Technology, June 1940.
- (13) H. O'Neill, *l. c.*, p. 10.
- (14) Glazebrook's 'Dictionary of Applied Physics,' vol. i. p. 392 (1922).
- (15) J. J. Bickerman and E. K. Rideal, *Phil. Mag.* (7) xxvii. p. 687 (1939).
- (16) W. B. Hardy and I. Doubleday, *Proc. Roy. Soc. A*, c. p. 550 (1922).
- (17) W. B. Hardy, *Phil. Trans. A*, cccxxx. p. 1 (1931).
- (18) R. Schnurmann, *J. Appl. Physics*, xi. p. 624 (1940).
- (19) J. M. Macaulay, 'Nature,' cxix. p. 425 (1927).
- (20) F. P. Bowden and D. Tabor, *Proc. Roy. Soc. A*, clxix. p. 391 (1939).
- (21) H. Blok, *Proceedings of the Conference on Friction*, Mass. Inst. Techn., June 1940.
- (22) G. Schmaltz, *Technische Oberflaechenkunde*, Berlin, 1936.

V. *Cloud and Earth Lightning Flashes.*

By J. ALAN CHALMERS, M.A., Ph.D.,
Lecturer in Physics in the Durham Colleges of
the University of Durham *.

[Received April 22, 1941.]

1. *Introduction.*

It is well known that thunderstorms in tropical and semi-tropical climates (*e. g.*, India and South Africa) do not cause as much damage as their frequency of occurrence might lead one to expect, and that this is due to the lightning flashes occurring mainly between the two poles of a cloud, rather than from the lower pole to the earth. On the other hand, in England and other temperate climates, flashes to earth are more frequent in comparison with cloud flashes.

Also, we often have, in winter in England, conditions which are similar meteorologically to those existing during thunderstorms, but without giving rise to lightning flashes; and Wormell ⁽¹⁾ has pointed out that point discharge currents in winter are often as great as those during summer thunderstorms, although there are no lightning discharges.

It appears that no attempt has hitherto been made to give any account of these facts, and it is the purpose of this paper to make a first advance in this direction, on the supposition that the differences depend on differences in temperature levels and cloud levels.

2. *Theories and Assumptions.*

It has been established by Simpson and Scrase ⁽²⁾ and Simpson and Robinson ⁽³⁾ that the electrical structure of a thundercloud consists mainly of a negative charge centred near the 0° isothermal, with a positive charge in the upper part of the cloud. This confirms the conclusions reached by Schonland and Craib ⁽⁴⁾ from the observations of field changes accompanying lightning flashes. In addition, Simpson and Scrase found evidence for the existence of a lower positive charge, at a temperature well above the freezing-point; this does not appear to be required to account for Schonland's observations.

We shall neglect the lower positive charge, if any, and we shall just consider the process of separation which gives rise to the negative charge

* Communicated by the Author.

at about the 0° C. level and the positive charge higher up. Simpson and Scrase ⁽²⁾ have suggested that the process is that of friction between ice-particles, the larger particles acquiring a negative charge and the smaller particles or the air acquiring a positive charge ; this agrees with the fact that the separation is found to occur at points higher up than the freezing-point level. Wormell ⁽⁵⁾ has suggested that the process may be that of the selective capture of ions, as suggested by Wilson ⁽⁶⁾, but with ice-particles rather than water drops as the agents.

In order to simplify the discussion, it will be assumed that the cloud can be taken to be of considerable horizontal extent, so that, mathematically, the problem is one-dimensional. It will be assumed that the occurrence of a lightning flash is conditioned by there being a sufficient potential gradient at some point ; a discharge will then start at that point and be propagated by the mechanism discussed by Schonland ⁽⁷⁾ and his co-workers. If the discharge starts above the greatest concentration of negative charge, then it will proceed to the upper pole of the cloud, if below then to the earth. Thus the relative frequencies of cloud and earth flashes can be discussed by discussing the potential gradients above and below the negative charge, and the possibility of similar conditions without discharge occurs when neither potential gradient is high enough to cause discharge.

It must be realised that the picture of a cloud of large horizontal extent is far removed from the actual state of a thundercloud, and that the charges in thunderclouds are confined within fairly narrow limits ; thus the potential gradient required to give sparking will occur, not over a wide area, but in a narrow region only. However, it seems that the one-dimensional picture can give useful results, for the concentration of lines of force from a wider area on to a more confined space will occur both above and below the negative charge and will alter the potential gradients in both cases. Comparing clouds which are, in general, similar but at different heights, the effect of the concentration of lines of force will be similar in different cases, and the ratio of the potential gradients in the one-dimensional case will give an indication as to the actual probabilities of lightning flashes in the real storm.

In order to make comparisons between storms under different conditions of temperature and humidity, we must make some assumptions ; we can assume that we shall compare storms in which the vertical air currents giving rise to the storms are similar. Since the total heights of thunderclouds are large compared with the heights of their bases or of their freezing-point levels, it is a sufficient first approximation to say that the distance between the upper positive and lower negative charge is nearly the same, whatever may be the height of the latter.

The air that rises in a thunderstorm, whether of "heat instability"

or "cold front" type, has been near the earth's surface for some time and its humidity determines a definite dew-point. The rising air falls in temperature according to the adiabatic law, and when it reaches the dew-point (slightly altered from the measured dew-point owing to alteration of vapour pressure with atmospheric pressure), condensation occurs and we have the base of the cloud; the air rising now leaves behind some of the condensed moisture and cools at a rate between the adiabatic and the "pseudoadiabatic," a rate which gives about half the cooling for the same change of height as compared with that below the cloud. At a certain height, the temperature has reached the freezing-point and ice crystals form; we need not consider the air at any higher level.

3. Potentials and Fields.

Wilson ⁽⁸⁾ has produced a theory of the variation of potential and potential gradient with height above the earth's surface, assuming that there is a constant unipolar current due to ions produced by point discharge, and assuming vertical flow of current, *i. e.*, a one-dimensional problem; the variation of potential gradient with height is due to the space charge of the positive ions which are moving upwards.

If X_1 and X_2 are the potential gradients at two points at a height difference h ,

$$X_2^2 - X_1^2 = \frac{8\pi i h}{w},$$

where i is the vertical ionic current density and w is the mobility of the ions concerned.

The potential difference ($V_2 - V_1$) is given by :

$$V_2 - V_1 = \left(\frac{32\pi i}{9w} \right)^{1/2} [(h+c)^{3/2} - c^{3/2}],$$

where

$$c = \frac{wX_1^2}{8\pi i}.$$

The same theory holds inside the cloud in the region between the base of the cloud and the concentration of negative charge, if we assume that the negatively charged ice-particles do not fall below this region; but, since the small ions, which carry the charge below the cloud will become attached to cloud-particles in the cloud, we must consider large ions, of mobility w' , rather than small ions of mobility w .

The boundary conditions at the base of the cloud are needed in order to obtain the constants of integration for Wilson's theory inside the cloud. We can assume the same vertical ionic current inside the cloud as outside, otherwise we should not get steady conditions, and the potential at the base of the cloud must be the same inside as outside the cloud. Unless there is a concentration of ions on the lower edge of the cloud, there can

be no change in potential gradient at the base of the cloud (for this would require a large value of $\frac{d^2V}{dx^2}$, which is proportional to the charge density) if the charges are all attached to particles of the size of large ions and none to larger particles, then there can be no concentration of large ions on the actual boundary, for such ions would move in the field. Thus we can take the potential gradient just inside the cloud to be the same as that just outside.

Compared with the other potential gradients, that at the earth's surface can be neglected, and, taking the potential of the earth to be zero, we get for the heights b , of the base of the cloud and z , of the freezing-point level :

$$X_b^2 = \frac{8\pi i b}{w},$$

$$X_z^2 - X_b^2 = \frac{8\pi i (z-b)}{w'},$$

$$X_z^2 = 8\pi i \left[\frac{(z-b)}{w'} + \frac{b}{w} \right],$$

$$V_b = \left(\frac{32\pi i}{9w} \right)^{1/2} b^{3/2},$$

$$V_z - V_b = \left(\frac{32\pi i}{9w'} \right)^{1/2} [(z-b+a)^{3/2} - a^{3/2}],$$

where $a = \frac{w' X_b^2}{8\pi i} = \frac{w' b}{w},$

$$V_z = \left(\frac{32\pi i}{9w'} \right)^{1/2} \left[\left(z-b + \frac{w' b}{w} \right)^{3/2} - \left(\frac{w' b}{w} \right)^{3/2} + \left(\frac{w'}{w} \right)^{1/2} b^{3/2} \right].$$

It is not possible to state definitely what is the potential gradient above the concentration of negative charge, but it clearly depends on V_z and is greater the greater is V_z .

Let us put $\frac{w'}{w} = \beta$ and $\frac{z-b}{b} = \alpha.$

Then $X_z^2 = \frac{8\pi i b}{w'} [\alpha + \beta],$

$$V_z = \left(\frac{32\pi i}{9w'} \right)^{1/2} b^{3/2} [\alpha + \beta]^{3/2} - \beta^{3/2} + \beta^{1/2}].$$

Giving $\frac{V_z}{X_z} = \frac{2b}{3} \frac{[(\alpha + \beta)^{3/2} - \beta^{3/2} + \beta^{1/2}]}{(\alpha + \beta)^{1/2}}.$

The ratio $w'/w=\beta$ can be taken as 1/500, and we find that, for constant b , $\frac{V_z}{\bar{X}_z}$ gives a minimum value for $\alpha=.08$ approximately.

If the temperature of the air before rising is T , and the dew-point is t , in Centigrade degrees, if the adiabatic fall in temperature is L° per cm. and, if inside the cloud it is approximately $\frac{1}{2}L^\circ$ per cm., we get

$$L=T-t \quad \text{and} \quad (z-b)\frac{L}{2}=t \text{ approximately ;}$$

then
$$\alpha = \frac{z-b}{b} = \frac{2t}{(T-t)},$$

and
$$\frac{V_z}{\bar{X}_z} = \frac{2(T-t)}{3L} \left[\frac{2t}{(T-t)} + \beta + \frac{\beta^{1/2}(1-\beta)}{\left(\frac{2t}{(T-t)} + \beta\right)^{1/2}} \right].$$

We must now consider two separate cases: in the first place, if the freezing-point is reached before the dew-point, then the concentration of negative charge is at the base of the cloud and we have to find $\frac{V_b}{\bar{X}_b}$, which is

$$\frac{2}{3}b \quad \text{or} \quad \frac{2}{3}\bar{L}, \text{ since } t \text{ is zero.}$$

In the other case, when z is greater than b , or t greater than zero,

$$\frac{V_z}{\bar{X}_z} = \frac{2}{3L} \left[2t + \beta(T-t) + \frac{(T-t)^{3/2}\beta^{1/2}(1-\beta)}{(2t + \beta(T-t))^{1/2}} \right].$$

Or, neglecting β in comparison with unity, approximately

$$\frac{V_z}{\bar{X}_z} = \frac{2}{3L} \left[2t + \beta T + \frac{(T-t)^{3/2}\beta^{1/2}}{(2t + \beta T)^{1/2}} \right].$$

Unless t is small compared with T , this gives approximately

$$\frac{V_z}{\bar{X}_z} = \frac{4t}{3L}.$$

When t is small compared with T , we get

$$\frac{V_z}{\bar{X}_z} = \frac{1}{3L} \frac{[(2t + \beta T)^{3/2} + T^{3/2}\beta^{1/2}]}{(2t + \beta T)^{1/2}},$$

giving a minimum value when $\frac{t}{T} = \left(\frac{\beta}{2}\right)^{1/3} = \frac{1}{25}$ approximately, the value being about $\frac{T}{20L}$.

4. Discussion.

The above results show that, unless t is small, *i.e.*, the dew-point is close to the freezing-point, the ratio V_z/X_z is proportional to t ; this

means that, irrespective of the ground temperature, the higher the dew-point temperature is, the more probable is a lightning flash within the cloud, in comparison with one to earth. This gives an account of the facts in regard to tropical thunderstorms, in which the dew-point temperature will certainly be high.

As t decreases, so will V_z/X_z , and hence we have the results that, in summer storms in England, earth flashes are more frequent.

It is possible to make predictions in other cases ; if we have very dry air, so that the dew-point is below the freezing-point, then V_z/X_z is even greater than for the case of t large, for it is $\frac{2T}{3L}$ as compared with $\frac{t}{3L}$;

in this case, then, we should expect a predominance of cloud flashes, this becoming less marked as the ground temperature decreases. This may be the case of South African storms.

If we now turn to the actual values of V_z and X_z , instead of their ratio, we see that, for any one value of i , both quantities increase as α increases, *i. e.*, t increases. Hence, for a given value of i , the sparking potential will be reached more readily the higher is the dew-point, suggesting that tropical storms give more flashes, not only because of greater electrical effects inside the cloud, but because of the high dew-point. We can say that the greater the height within the cloud at which the freezing-point occurs, the greater will be fields and potentials, and so the sooner, in the process of separation, will lightning flashes occur.

If V_z and X_z are both smaller than the values required for a lightning flash while i , the point discharge current density is equal to the rate of separation of charge within the cloud, so that the conditions are steady, then there will be no lightning, although meteorological conditions may be similar to those in a thunderstorm. This is more likely to occur the lower is the dew-point temperature, and if the dew-point is below the freezing-point, then the effect is most likely when " b ," the height of the base of the cloud, is low, *i. e.*, in cold humid air, exactly the conditions of the cases in English winter, when lightning flashes are rare.

It must be realized that the present discussion can be little more than the first approach to the subject, for it would appear from the above results that, under certain conditions, there must be either cloud flashes or earth flashes, but not both except in an exceptional case, with a particular value of V_z/X_z . The fact that in the same storm, without much alteration of conditions, there may be both types of flash, is accounted for by the fact that the above analysis does not consider the actual building up of the charges in the cloud, and the possibility that, after a flash in one direction, there might be enough potential gradient remaining at another point for a flash to start there more readily than for the building up of charge to give another flash similar to the first.

There are many other respects in which the present analysis is incomplete, a number of factors such as the charges carried by rain and the convection currents due to the vertical motion of the air. But the results appear to be sufficient to show that the factors taken into consideration may be the most important in determining the direction of the discharge.

References.

- (1) Wormell, T. W., Proc. Roy. Soc. cxv. p. 443 (1927).
- (2) Simpson, Sir G. C., and Scrase, F. J., Proc. Roy. Soc. clxi. p. 309 (1937).
- (3) Simpson, Sir G. C., and Robinson, G. D., Proc. Roy. Soc. clxxvii. p. 281 (1941).
- (4) Schonland, B. F. J., and Craib, J., Proc. Roy. Soc. cxiv. p. 229 (1927).
- (5) Wormell, T. W., Phil. Trans. Roy. Soc. ccxxxviii. p. 249 (1939).
- (6) Wilson, C. T. R., J. Frank. Inst. ccviii. p. 1 (1929).
- (7) Schonland, B. F. J., Proc. Roy. Soc. clxiv. p. 132 (1938).
- (8) Wilson, C. T. R., Proc. Phys. Soc. xxxvii. p. 32 D (1925).

VI. *On some Expansions containing Laguerre Polynomials,
and their Expressions in Terms of Whittaker's
Confluent Hypergeometric Functions.*

By D. P. BANERJEE, Mymensingh (India) *.

[Received May 31, 1940.]

THE Laguerre Polynomials are defined by the identity †

$$(1-t)^{-\alpha-1} \exp\left(\frac{xt}{t-1}\right) \equiv \sum_{n=0}^{\infty} t^n L_n^{\alpha}(x). \quad (1)$$

This gives

$$\begin{aligned} \sum_{n=0}^{\infty} t^n L_n^{\alpha+\beta+1}(x+y) &= (1-t)^{-\alpha-\beta-2} \exp\left\{\frac{(x+y)t}{t-1}\right\} \\ &= (1-t)^{-\alpha-1} \exp\left(\frac{xt}{t-1}\right) \times (1-t)^{-\beta-1} \exp\left(\frac{yt}{t-1}\right) \\ &= \sum_{r=0}^{\infty} t^r L_r^{\alpha}(x) \sum_{s=0}^{\infty} t^s L_s^{\beta}(y). \end{aligned}$$

On equating the coefficients of t^n on both sides of this identity, it is found that

$$L_n^{\alpha+\beta+1}(x+y) = \sum_{s=0}^n L_{n-s}^{\alpha}(x) L_s^{\beta}(y). \quad (2)$$

Again, from (1),

$$\begin{aligned} \sum_{n=0}^{\infty} t^n L_n^{\alpha}(x+y) &= (1-t)^{-\alpha-1} \exp\left\{\frac{(x+y)t}{t-1}\right\} \\ &= \sum_{r=0}^{\infty} \frac{(-yt)^r}{r!} (1-t)^{-\alpha-r-1} \exp\left(\frac{xt}{t-1}\right) \\ &= \sum_{r=0}^{\infty} \frac{(-y)^r}{r!} t^r \sum_{s=0}^{\infty} t^s L_s^{\alpha+r}(x). \end{aligned}$$

Hence, equating the coefficients of t^n , we have

$$L_n^{\alpha}(x+y) = \sum_{r=0}^n \frac{(-y)^r}{r!} L_{n-r}^{\alpha+r}(x). \quad (3)$$

* Communicated by the Author.

† McArthur, Proc. Edin. Math. Soc. xxxviii. p. 27 (1920).

These formulæ can be expressed in terms of confluent Hypergeometric Functions by means of the relation

$$L_n^\alpha(x) = \frac{(-1)^n}{n!} x^{-\frac{1}{2}\alpha - \frac{1}{2}} e^{\frac{1}{2}x} W_{n+\frac{1}{2}\alpha+\frac{1}{2}, \frac{1}{2}\alpha}(x), \quad \dots \quad (4)$$

where $n=0, 1, 2, 3, \dots$

Thus (2) reduces to

$$\begin{aligned} W_{n+\frac{1}{2}\alpha+\frac{1}{2}\beta+1, \frac{1}{2}\alpha+\frac{1}{2}\beta+\frac{1}{2}}(x+y) &= (x+y)^{\frac{1}{2}\alpha+\frac{1}{2}\beta+1} x^{-\frac{1}{2}\alpha-\frac{1}{2}} y^{-\frac{1}{2}\beta-\frac{1}{2}} \\ &\times \sum_{s=0}^n {}^nC_s W_{n-s+\frac{1}{2}\alpha+\frac{1}{2}, \frac{1}{2}\alpha}(x) W_{s+\frac{1}{2}\beta+\frac{1}{2}, \frac{1}{2}\beta}(y). \quad \dots \quad (5) \end{aligned}$$

Next, (3) gives

$$\begin{aligned} W_{n+\frac{1}{2}\alpha+\frac{1}{2}, \frac{1}{2}\alpha}(x+y) &= e^{-\frac{1}{2}y} (x+y)^{\frac{1}{2}\alpha+\frac{1}{2}} \\ &\times \sum_{r=0}^n {}^nC_r y^r x^{-\frac{1}{2}\alpha-\frac{1}{2}r-\frac{1}{2}} W_{n+\frac{1}{2}\alpha-\frac{1}{2}r+\frac{1}{2}, \frac{1}{2}\alpha+\frac{1}{2}r}(x). \end{aligned}$$

Again, substituting from (4) into (1) and putting $2\alpha-1$ for α , we get

$$x^\alpha (1-t)^{-2\alpha} \exp\left\{\frac{x(t+1)}{2(t-1)}\right\} = \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} t^n W_{n+\alpha, \alpha-\frac{1}{2}}(x). \quad \dots \quad (6)$$

Here put $x+y$ for x and m for α , where m is a positive integer. Then

$$\begin{aligned} \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} t^n W_{n+m, m-\frac{1}{2}}(x+y) &= (x+y)^m (1-t)^{-2m} \exp\left\{\frac{x(t+1)}{2(t-1)}\right\} \exp\left\{\frac{y(t+1)}{2(t-1)}\right\} \\ &= \sum_{p=0}^m {}^mC_p \sum_{r=0}^{\infty} \frac{(-t)^r}{r!} W_{r+m-p, m-p-\frac{1}{2}}(x) \\ &\times \sum_{s=0}^{\infty} \frac{(-t)^s}{s!} W_{s+p, p-\frac{1}{2}}(y). \quad \dots \quad (7) \end{aligned}$$

Hence, on equating the coefficients of t^n , we have

$$\begin{aligned} W_{n+m, m-\frac{1}{2}}(x+y) &= \sum_{p=0}^m {}^mC_p \sum_{r=0}^n {}^nC_r W_{n-r+m-p, m-p-\frac{1}{2}}(x) W_{r+p, p-\frac{1}{2}}(y). \quad \dots \quad (8) \end{aligned}$$

VII. *Notices respecting New Books.*

Introduction to Algebraic Theories. By A. ADRIAN ALBERT. [Pp. viii+137.] (U.S.A. : The University of Chicago Press; Great Britain and Ireland : The Cambridge University Press, 1941. Price 10s. 6d.)

PROFESSOR ALBERT has written this book as a bridge between the elementary treatises on algebra and the rigorous abstract treatment which has become possible owing to the thorough examination of fundamentals undertaken in recent years, and which, because of its simplicity and completeness, is necessarily adopted in any advanced treatment.

There are six chapters, the first three of which deal respectively with polynomials (mainly in one variable), matrices, and the equivalence of matrices and of forms. Towards the end of this chapter, the concept of a (non-modular) field is first introduced. The first three chapters should make very easy reading, even for beginners at the subject. In the remaining three, modern abstract concepts crowd much closer together. In the course of chapter 4, linear spaces are defined and discussed, together with such refined concepts as that of linear mapping. Chapter 5, on polynomials with matrix coefficients, perhaps makes greater demands on technique than the others, and the final chapter, headed *Fundamental Concepts* deals with groups, rings, integral domains, ideals and quadratic fields. It should thus fit the student to tackle either the more advanced treatises, or even original papers.

Topological Groups. By L. PONTRJAGIN, translated by EMMA LEHMER. [Pp. ix +299.] (U.S.A. : Princeton University Press; Great Britain : Sir Humphrey Milford, at the Oxford University Press, 1939. Price 4\$.)

THIS, the second volume in the Princeton Mathematical Series, was originally written in Russian. The subject is one opened up by the author himself, who explains that he regards topological spaces as the study of the pure operation of passage to a limit, just as the theory of groups studies multiplication in its purest form. The topological group results from a combination of the two ideas.

The matter is set out in formal detail, with sequences of numbered theorems, but is very clearly expressed (a tribute both to author and to translator) and forms a very readable exposition of the whole theory.

The Theory and Applications of Harmonic Integrals. By W. V. D. HODGE. (Cambridge, at the University Press; 1941. Price 21s. net.)

THIS volume is based on an Adam's Prize essay. It deals with integrals on a manifold of a class of forms the definition of which has analogies with the definition of potential functions in mathematical physics or of harmonic functions in the theory of functions of a complex variable. Most of the material has, so far, only been available in scattered papers, many of them written by Professor Hodge himself. The publication of this treatise should therefore be of great service to students and research workers in the subject.

The book is divided into five long chapters. The first contains a readable summary of the tensor calculus and of the topology of Riemannian manifolds. The second deals with the properties of multiple integrals on a manifold, while in the third harmonic integrals are defined and their properties established. The two remaining chapters are devoted to the applications of these integrals, chapter IV. to the properties of harmonic integrals in the Riemannian of an algebraic variety and chapter V. to the application of the theory of harmonic integrals to certain problems in the theory of continuous groups. Bibliographies of relevant books and papers are provided at the ends of the chapters.

This treatise is obviously intended for specialists ; it should prove an incentive to further original investigations in the geometry of n dimensions.

The Physical Society : Reports on Progress in Physics.—Volume VII. (1940) ; General Editor, J. H. AWBERY. [Pp. 362.] (Price to non-Fellows, 22s. 6d. net. London, 1941.)

THE new volume shows a little reduction in size, as compared with its immediate predecessors. It returns, in fact, approximately to the dimensions of volumes I. and II.—1938 and 1939 were peak years in point of size. It demands no great imaginative or reflective powers to find reasons for the slight falling-off ; in fact, considering the present situation—or even ignoring it entirely—the production of this volume must be regarded as a most creditable performance. It need hardly be said that there is no corresponding falling-off in the quality of the articles.

The general plan of these Reports is well known. This volume contains eighteen separate articles, which cannot here be described in detail. E. G. Richardson writes on sound and on musical acoustics, W. D. Wright on photo-electric spectrophotometry and its industrial applications, H. H. Nielsen on infra-red spectra of polyatomic molecules, N. Feather on gamma radiations emitted in nuclear processes, R. Peierls on the Bohr theory of nuclear reaction, O. Klemperer on electron microscopes of high magnification, H. W. Lee on new lens systems, A. Hunter on absorption of light in interstellar space, A. D. Thackeray on solar physics, R. C. Brown on surface tension, J. A. Beattie and W. H. Stockmayer on equations of state. By far the longest article (73 pages) is that on electrical discharges in gases and their applications, contributed by V. J. Francis and H. G. Jenkins, of the G.E.C. The second part of this, which is to deal with the practical applications, is provisionally promised for the next volume of *Reports*. J. K. Roberts writes on interactions of gases with metals and crystalline solids, F. Eirich on viscosity and the nature of substances of high molecular weight in solution, and finally, the first Editor of the *Progress Reports*, Prof. Allan Ferguson, writing on "the development of the teaching of experimental physics in British universities," continues the series of "teaching" articles started in earlier volumes by Barton and Owen.

It is always a little humiliating for a reviewer to be reduced to the mere enumeration of chapter headings ; if, however, we are prepared to accept as a reasonable, if somewhat novel, assumption that one of his main functions is to let the prospective reader see what he is being offered, there is every excuse in the present instance for following so dull a plan. All the articles are good, and some of them are very good, and they conform to the standards which have now become traditional in the *Progress* volumes. It might be said with some justice that this volume caters rather more for specialists, and rather less for physicists of more diffused interests, than some of its predecessors, but it would

be an odd physicist who could not find a great deal both of profit and enjoyment in its pages.

The bibliographies follow the new-fangled plan of listing all references alphabetically under author's names, with dates, instead of numbering them consecutively through each article. The older plan often served as an index to references otherwise buried without trace in a long article, but no doubt there is some good reason for the change; at all events, in making it the *Reports* are following a practice initiated by reputable scientific journals, and presumably it has not been done without consideration. It is, however, mildly disliked by at least one reader.

This is, however, a trivial grumble to make in the face of so much excellence. Perhaps a reviewer who is unable to make a display of erudition may be forgiven for taking, in compensation, an opportunity for the exhibition of his ingrained conservatism. The volume really merits unqualified commendation.

H. R. R.

A Text-book of Sound. By Dr. A. B. WOOD. [Pp. xvi+578.] (London: G. Bell & Sons, Ltd., 1941. Price 30s. net.)

To those who were educated in the atmosphere of the text-books of twenty or thirty years ago, Dr. Wood's treatise came as a very welcome change. Few subjects which may still be treated by classical methods have shown such a striking degree of development as has been evident in the science of Acoustics—or, as Dr. Wood prefers it, of Sound, for he reserves the term "Acoustics" for phenomena within the range of hearing. The problem of noise, the great increase in the range of available frequencies, the analogy between electrical and mechanical vibrating systems, the acoustics of buildings, echo-sounding and sound-ranging—these are but a few of the many topics discussed by the author. It must not, however, be imagined that discussions of technical applications dominate the treatment. Far from it; the treatise is a logical and full development of the theory and practice of Sound. One section of the book is devoted to technical applications; the four remaining sections are concerned with the theory of vibration, vibrating systems and sources of sound, the transmission of sound, and the reception and measurement of sound energy.

The first edition appeared in 1930 and was reprinted in 1932 and in 1937. In this new and enlarged edition the author has incorporated accounts of recent advances and has provided many new references to the literature of the subject. He is to be congratulated on having preserved the essential unity of outlook of the first edition.

A. F.

An Introduction to Astronomy. By Prof. R. H. BAKER. [Pp. 315.] (New York: D. van Nostrand Co. Inc.; London: Macmillan & Co., 1940. Price 12s. 6d. net.)

PROFESSOR BAKER's larger text-book is well known to all students of astronomical science. This simpler introduction to astronomy is a non-mathematical and non-technical study which, in its present form (the second edition), begins with the very elements of the subject and provides the reader with brief, clear and authoritative surveys of the latest advances in the science.

It is designed as a one-semester course for beginners in astronomy; the general reader will find in it an admirable and up-to-date introduction to the subject.

A. F.

[The Editors do not hold themselves responsible for the views expressed by their correspondents.]

VIII. *A Study of the Growth of Chemical Compounds on a Copper Single Crystal by the Method of Electron Diffraction.*

By I. H. USMANI, B.Sc., Ph.D., D.I.C.(Lond.).*

[Plates I.-III.]

[Received February 7, 1941.]

1. *Introduction.*

The oriented overgrowth of one crystal on another is well known and has been studied both by X-rays and electron diffraction. Using the latter method, G. P. Thomson (1931) showed that a copper compound is oriented in a definite way on the etched surface of a copper single crystal. Later, Cochrane (1936) deposited a number of metals electrolytically on an etched face of a copper crystal and found by electron diffraction that metals with a cubic structure could grow as single crystals on a copper crystal base at a certain current density. Lassen (1934), Brück (1936), and Lassen and Brück (1935) have examined the structures of thin metal films evaporated *in vacuo* on the cube face of a rock salt and found that there is complete orientation of the deposit at a certain critical temperature of the rock salt. Their work was further extended by Rüdiger (1937), who evaporated gold, silver, and palladium on to the cleavage faces of mica, calcite, and fluorite. The dependence of orientation on temperature of the base was confirmed. A notable feature of these experiments is the law, generally known as Brück's law, governing the orientation of the deposits. It was found that in many cases orientation was such as to make the distances between the nearest atoms of the crystal and the base a minimum, reckoned over the first cell.

Recently, however, crystals have been attacked chemically and the resulting film deposit on the surface examined by electron diffraction. Thus, Yamaguti (1938) heated an etched copper crystal to high temperatures and found that the oxide film formed showed two orientations at two different temperatures. Miyaki (1938) studied the growth of Sb_2O_3 on a crystal of Sb_2S_3 by heating the latter to high temperatures, and observed single crystal orientation of the oxide.

The present paper describes the work undertaken to study the growth of films formed by chemical attack on the etched face of a copper single

* Communicated by Prof. G. P. Thomson.

crystal. The reflexion method of electron diffraction was mostly used for photographing patterns from the surfaces under examination. However, when the surface film of the compound could be stripped from the copper crystal base it was examined by transmission.

2. General Experimental Procedure.

The electron diffraction camera used was of Thomson and Fraser (1930) type with slight modifications. The crystal specimens were cut from a cylindrical rod of copper crystal, about 6 mm. in diameter, to any desired face by sawing and grinding. A specimen thus obtained was ground on 0000 emery paper under benzene, washed in acetone and then water, to be finally etched by 10 per cent. ammonium persulphate solution for five to fifteen minutes. The etched surface was washed under running water and always examined in the camera before subjecting it to any chemical treatment. The usual cross-grating patterns of spots were obtained, rich whenever an important zone axis of the crystal was parallel to the incident electron beam.

G. P. Thomson (1931) has shown that these spots are due to small projections on the surface of the metallic crystal brought out by etching, and that the absence of Kikuchi lines from such crystals is due to the slight bending and distortion caused during the cutting and grinding operations.

3. Cuprous Iodide on Copper.

A specimen of copper crystal was cut having its surface parallel to the (111) plane. It was put in the diffraction camera after the etching treatment and examined in the two important azimuths of this plane viz., $\bar{1}\bar{1}2$ and $1\bar{1}0$ at right angles to one another. The two diffraction patterns in the above two azimuths are shown in figs. 1 and 2 (Pl. I.) respectively.

The crystal was taken out of the camera and hung as an anode in an electrolytic bath with the following composition :—

Sodium iodide	1 gm.
Distilled water	100 c.c.

Usually, when the current density was high, a deposit was formed on the copper crystal as cuprous iodide ; but it always yielded Debye-Scherrer rings characteristic of a polycrystalline film on examination in the camera. At very low current density either nothing was formed or a deposit of cuprous iodide also in the polycrystalline form. By trial and error, however, a suitable range of current density (between 0.28 ma./cm.² and 0.819 ma./cm.²) was found, which gave a deposit of cuprous iodide in the form of a single crystal, shown by the appearance of spots instead of rings in the subsequent photographs. One such photograph, taken

from a cuprous iodide film deposited at a current density of 0.819 ma./cm.² for 30 sec., is shown in fig. 3 (Pl. I.). Besides the rings, which are due to polycrystalline cuprous iodide, spots due to the copper crystal base as well as due to the cuprous iodide film, in the form of a single mosaic crystal, are apparent.

To study the growth of this film by degrees, the time of treatment was increased to five minutes at the same current density. The film thus prepared was then examined in the camera in such a way that the copper crystal was in its $1\bar{1}0$ azimuth, as indicated by the marks on the manipulating greased cone outside. One of the photographs taken in this position is shown in fig. 4 (Pl. I.). It can be seen from this photograph that although some polycrystalline cuprous iodide is present, giving rise to rings, much of it is in the form of a single crystal giving prominent spots on the photographic plate. The cross-grating nature of the spots is almost complete. The film has grown to a thickness sufficient to obscure all the spots due to the copper base.

When, however, this deposit was examined in the 112 azimuth of the copper base, rows of spots were visible, all of them occurring in pairs. They can be seen in fig. 5 (Pl. II.). It was found that further increase in the time of treatment at the above critical current density did not help. Patterns similar to those shown in fig. 4 (Pl. I.) and fig. 5 (Pl. II.) continued repeating up to fifteen minutes of treatment of the crystal in the electrolytic bath. After that time the copper base lost its influence and the top layers degenerated completely into a polycrystalline deposit.

4. Theory.

Diffraction patterns, and especially those of electron waves, are very conveniently interpreted in terms of Ewald's (1921) reciprocal lattice.

If O be the origin of a unit cell of a simple crystal lattice defined by the three elementary vectors $\vec{\alpha}_1$, $\vec{\alpha}_2$, and $\vec{\alpha}_3$, then any lattice point R is given by the vector $\vec{r} = n_1\vec{\alpha}_1 + n_2\vec{\alpha}_2 + n_3\vec{\alpha}_3$, where n_1 , n_2 , and n_3 are integers. Let there be three vectors $\vec{\beta}_1$, $\vec{\beta}_2$, and $\vec{\beta}_3$ such that $\vec{\beta}_1$ is perpendicular to the plane containing $\vec{\alpha}_2$ and $\vec{\alpha}_3$ and has a magnitude $[\alpha_2 \cdot \alpha_3]/v$, where the brackets [] stand for the vector product and v is the volume of the unit cell. The scalar magnitude of this vector, by definition of "vector product" would therefore be $\vec{\alpha}_2 \cdot \vec{\alpha}_3 \cdot \sin(\alpha_2, \alpha_3)/v$.

Similarly
$$\vec{\beta}_2 = [\alpha_1 \cdot \alpha_3]/v$$
 and
$$\vec{\beta}_3 = [\alpha_1 \cdot \alpha_2]/v.$$

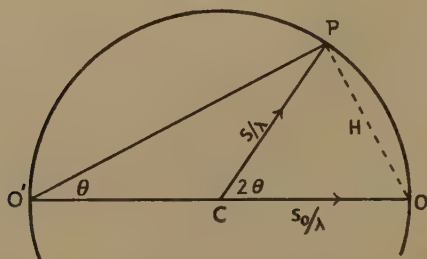
The vectors $\vec{\beta}_1$, $\vec{\beta}_2$, and $\vec{\beta}_3$ are the vectors of a new lattice called the reciprocal lattice. Any point of the reciprocal lattice is consequently

given by a vector $\vec{H} = A_1\vec{\beta}_1 + A_2\vec{\beta}_2 + A_3\vec{\beta}_3$, where A_1, A_2 , and A_3 are numerical coefficients.

The connexion between the reciprocal lattice and the law of diffraction can easily be seen with reference to fig. 6. Let $OC = 1/\lambda$ be drawn from O , the origin of the reciprocal lattice, in a direction opposite to that of the incident beam of electrons of wave-length λ . With C as centre and radius OC draw a sphere to meet OC produced in O' . It can be shown that if this sphere passes through any lattice point P of the reciprocal lattice, CP would be the direction of the diffracted beam obeying Laue and Bragg conditions.

For electrons of about 40 kv. used in practice the wave-length is 0.06 Å. ($\lambda = h/mv$), which gives the radius of Ewald's sphere as 16.7^{-1} Å. Now the distance between the cube corners of a copper crystal, for example, is nearly 0.55^{-1} Å. In approximation, therefore, electron diffraction pattern

Fig. 6.



can be interpreted in terms of the plane sections of the reciprocal lattice, taken at right angles to the incident beam.

From what has already been said about the reciprocal lattice, it is obvious that the reciprocal of a simple cubic lattice of edge a would be another cubic lattice of edge $1/a$. If, however, the atoms also occupy the centres of the faces of the simple cube, as is the case for copper and most of the compounds with which we are concerned, the points of the reciprocal lattice do not lie on a face-centred cubic framework, but on a body-centred cubic one of side $2/a$.

A plane section of such a lattice at right angles to the direction 112 will give points arranged in rectangles; while a section at right angles to 110 would give a "diamond" arrangement of points. Photographs shown in figs. 1 and 2 (Pl. I.) are thus accounted for. Indexing of the spots (by using the formula $d = L\lambda/2$) in fig. 1 also shows that they come from planes which have 112 as the common zone axis, while those in fig. 2 belong to the zone axis 110.

5. Orientation Relationship.

The X-ray data on the crystal structure of cuprous iodide reveal that it has a cubic structure of the ZnS type with an edge = 6.05 Å. The coordinates of the copper atoms are (0 0 0), $(0 \frac{1}{2} \frac{1}{2})$, $(\frac{1}{2} 0 \frac{1}{2})$ and $(\frac{1}{2} \frac{1}{2} 0)$ and those of iodine are $(\frac{1}{4} \frac{1}{4} \frac{1}{4})$, $(\frac{3}{4} \frac{3}{4} \frac{1}{4})$, $(\frac{3}{4} \frac{1}{4} \frac{3}{4})$, and $(\frac{1}{4} \frac{3}{4} \frac{3}{4})$. If the atomic scattering factor of copper be E_{cu} and that of iodine be E_{i} , the amplitude of the electron beam scattered by both the copper and iodine lattices would be given by

$$\begin{aligned}\psi &= \psi_{\text{cu}} + \psi_{\text{i}} \\ &= S \cdot \phi \cdot \{E_{\text{cu}} + E_{\text{i}} \cdot e^{\pi/2 i(A_1 + A_2 + A_3)}\},\end{aligned}$$

TABLE I.

Measurements of Plate of Fig. 3 (Pl. I.).

Current density : 0.819 ma./cm.²

Treatment time : 30 sec.

Camera length : 38.16 cm.

Wave-length : 0.0666 Å.

<i>r.</i>	<i>d.</i>	X-rays <i>d.</i>	Indices.
Cuprous iodide spots.			
cm.	Å.	Å.	
0.73	3.449	3.490	(111)
1.47	1.724	1.748	(222)
2.19	1.145	1.170	(333)
1.91	1.325	1.257	(420)
2.45	1.002	1.008	(531)
Copper spots.			
2.47	1.030	1.041	(222)
3.70	0.687	0.694	(333)
2.35	1.081	1.087	(113)
3.50	0.726	0.735	(224)
4.07	0.541	—	(335)

The above measurements give :—

Azimuth of copper crystal base	$\bar{1}\bar{1}0$
Azimuth of cuprous iodide	$\bar{1}\bar{1}2$
Plane of copper crystal parallel to free surface	(111)
Plane of cuprous iodide parallel to free surface	(111)

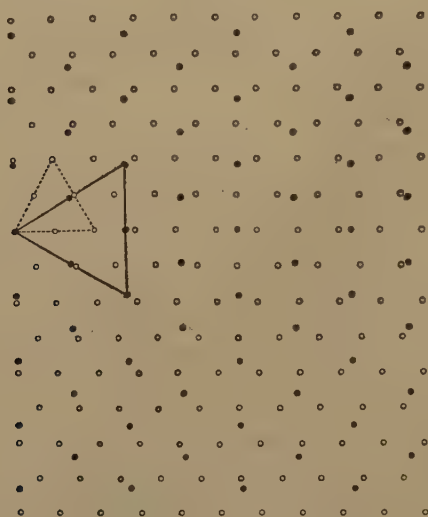
When S and ϕ are given by

$$\begin{aligned}\phi &= \sum e^{2\pi i(A_1x_1 + A_2x_2 + A_3x_3)}, \\ S &= \{1 + e^{\pi i(A_1 + A_2)} + e^{\pi i(A_2 + A_3)} + e^{\pi i(A_3 + A_1)}\}.\end{aligned}$$

From the second equation it would be seen that so long as E_{cu} and E are different, the points of the reciprocal lattice of cuprous iodide are the same as those allowed by the values of A 's, which are again either all even or odd. Thus patterns from cuprous iodide are plane sections of a body-centred cubic lattice of edge $2/6.05 \text{ \AA}$. at right angles to the electron beam.

The photograph in fig. 3 (Pl. I.) shows spots due to the copper crystal base arranged in "diamonds" and faint ones, lying on the rings due to the mosaic single crystals of cuprous iodide, arranged in "rectangles." Now the spots lying on the normal to the shadow edge in the photograph

Fig. 7.



are due to planes parallel to the free surface of the specimen, and measurements of the photographic plate tabulated in Table I. show that the (111) plane of copper crystal base (which was parallel to the free surface to start with) is parallel to the (111) plane of the cuprous iodide lattice. Thus it is clear that the cuprous iodide lattice grows on the copper crystal base in such a way that the (111) plane of cuprous iodide is parallel to the (111) plane of copper with the $\bar{1}\bar{1}2$ direction in the cuprous iodide lattice plane parallel to the $1\bar{1}0$ direction in copper. In other words, the two (111) planes are twisted in their own plane through an angle of 30° . The positions of copper atoms in the (111) plane of the copper crystal base and those in the (111) plane of the grown cuprous iodide lattice are shown in fig. 7 by the white and black circles respectively. The photograph

shown in fig. 4 (Pl. I.), was taken when the copper crystal was set in its $1\bar{1}0$ azimuth. The copper spots are not visible at all, due to a thicker deposit of cuprous iodide. As the film grows the spots due to cuprous iodide become sharper and well defined. Measurement of this plate gives the azimuth of cuprous iodide as $\bar{1}12$, again confirming the above orientation relationship.

Twinning.

The photograph in fig. 5 (Pl. II.), when regarded as a plane section of the reciprocal lattice, should show a "diamond" arrangement of points, for the azimuth of the copper crystal was 112 and that of cuprous iodide therefore $1\bar{1}0$. On close examination, however, it is found that the spots occur in pairs in lines on the photographic plate parallel to the shadow edge, and that each of these spots is due to cuprous iodide; in fact there are two "diamond" patterns superimposed on one another.

During crystal growth it often happens that two (or more) individual crystals of the same substance grow simultaneously, uniting together in a definite manner (other than parallelism) characteristic of the substance. In the case of only two crystals, *i. e.*, in the case of twins, the union is such that one face of each is in mutual contact. In the spinel type of twinning, where the octahedral planes are in mutual contact, to produce the twins the two individuals are first supposed to be set parallel to each other and then one of them is rotated with respect to the other through 180° about an axis called the twin axis. The twin plane is the contact plane, and hence the twin axis normal to it.

On analysing the plate in fig. 5 (Pl. II.) it is found that the spots of cuprous iodide lying on the normal to the shadow edge are due to reflexions from the (111) planes and do not occur in pairs. Consequently, if it is supposed that, during its growth, cuprous iodide film twins on the (111) plane (spinel twinning), the two "diamond" arrangements of spots on the plate could be assigned to two individual crystals of cuprous iodide. In fact, if A be a spot of one diamond arrangement due to one crystal, then its reflexion B across the normal to the shadow edge on the plate is a member of the other "diamond" arrangement due to the second individual.

Alternatively, reflexions of one pattern might be taken across a line parallel to the shadow edge to get the other pattern due to the second crystal. Similar twinning has been reported before by Cochrane (1936) in connexion with the study of the growth of electro-deposits on a copper single crystal. In our case the twin plane is fixed by the fact that the spots due to the (111) planes on the normal to the shadow edge did not occur in pairs and that twinning was not visible when the cuprous iodide

film was examined in the $\bar{1}\bar{1}2$ azimuth as shown by fig. 4 (Pl. I.). Although present, twinning would not show in this azimuth because the arrangement of atoms in the (111) planes, due to two individuals in contact with one another, is not different from the arrangement of atoms in the (111) plane of only one individual seen along the $1\bar{1}2$ direction. In other words, the beam distinguishes the twins along the $1\bar{1}0$ direction, but not along $\bar{1}\bar{1}2$; so that along the former direction patterns from two distinct individuals are obtained, while along the latter they superimpose to give a pattern as if it were due to one crystal.

6. Cuprous Bromide on Copper.

Copper crystal cut to (111) face was again used as a base. However, when sodium bromide was used instead of sodium iodide in the electrolytic solution, a deposit of cuprous oxide was invariably produced at all current densities. This was due to the alkaline tendency of the solution in the vicinity of the copper anode.

A few drops of sulphuric acid in the solution neutralized the alkali, and films of cuprous bromide could then be obtained at suitable current densities. The final composition of the solution used as a bath was as follows :—

Sodium bromide.....	1 gm.
Sulphuric acid	2.5 c.c.
Distilled water	100 c.c.

The addition of the acid made the solution very much more conducting, and the growth of cuprous bromide film in the form of a single crystal was seen to occur at a current density of 3.0 ma./cm.²

As before, photographs were taken from the deposited bromide film in various stages of its growth. Two such photographs are shown in figs. 8 and 9 (Pl. II.). They were exposed when the copper crystal base was set in $1\bar{1}0$ and $\bar{1}\bar{1}2$ azimuths respectively. They show exact similarity to those of fig. 4 (Pl. I) and fig. 5 (Pl. II.).

The crystal structure of cuprous bromide as determined by X-rays is also of the ZnS type, and consequently the explanations given for the patterns of cuprous iodide hold equally good for those of cuprous bromide. Spinel twinning is visible again in the $1\bar{1}0$ azimuth, and the measurement of the plates shows that the orientation relationship between the base and the deposit is the same as that found for cuprous iodide. This is shown in fig. 10.

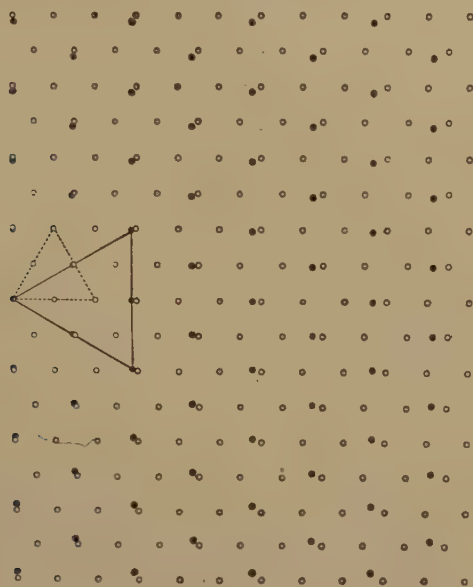
7. Cuprous Oxide on Copper.

During the process of deposition of cuprous bromide on the (111) face of copper crystal, an oxide film was invariably produced when the electrolyte was not acidified. Some of these oxide films when examined in the

camera revealed a marked orientation with respect to the copper substrate. Distinct spot patterns from such films are shown in fig. 11 (Pl. II.) and fig. 12 (Pl. III.), which were obtained when the copper crystal was set in $\bar{1}\bar{1}2$ and $\bar{1}\bar{1}0$ azimuths respectively. In both the photographs copper spots are visible besides those due to the oxide film.

The crystal structure of cuprous oxide is well known to be cubic, with an edge of 4.25 Å. The copper atoms lie on a face-centred cubic lattice

Fig. 10.



and the oxygen atoms on a body-centred lattice with the following coordinates in the unit cell :—

$$2\text{O} \dots\dots\dots (0\ 0\ 0), \text{ and } (\frac{1}{2}\ \frac{1}{2}\ \frac{1}{2}).$$

$$4\text{Cu} \dots\dots\dots (\frac{1}{4}\ \frac{1}{4}\ \frac{1}{4}), (\frac{1}{4}\ \frac{3}{4}\ \frac{3}{4}), (\frac{3}{4}\ \frac{1}{4}\ \frac{3}{4}), \text{ and } (\frac{3}{4}\ \frac{3}{4}\ \frac{1}{4}).$$

From the analysis of the patterns already mentioned it was found that the lattice of the oxide film follows the orientation of the copper base : (111) plane of the growing oxide being parallel to the (111) plane of the copper substrate such that the zone axes 112 and 110 of the one are parallel to the 112 and 110 axes of the other. In other words, the cube edges of the two crystals are parallel. It was also found that there were some spots, half-way between the origin and the allowed spots, which were forbidden by the structure factor of cuprous oxide, *e. g.* (100).

Although similar orientation was observed before by G. P. Thomson (1931), no forbidden spots were reported by him owing, perhaps, to their faint appearance. Subsequent experiments by Yamaguti (1938), however, on the formation of oxide films by heating a copper crystal at high temperatures, revealed the existence of three different orientations, and in some cases forbidden spots similar to the ones observed here. These three different orientations determined by him did not fix the relation of the oxide lattice to that of the copper base exactly. The plane of the oxide parallel to a plane of copper was not mentioned, which makes the orientation relationship indeterminate.

Fig. 13.



- and ○ represent allowed points in the section of the reciprocal lattice of Cu_2O at right angles to $1\bar{1}0$ direction.
 + represent points forbidden by the structure factor of Cu_2O .

To verify Yamaguti's observations, a specimen of copper crystal was put on a red hot tin plate to oxidize. The resulting oxide film gave patterns similar to those shown by the film formed electrolytically, revealing orientation of the same nature as already mentioned. The three different orientations observed by Yamaguti might perhaps be due to the recrystallization of the copper crystal at the high temperatures he used.

However, his explanation of the forbidden spots which he also observed applies to the ones mentioned above. Spots of this nature have been observed before by Raether (1932) from the surfaces of iron pyrites, and an analogous explanation was given.

To take only one azimuth, fig. 13 shows a section of the reciprocal lattice of cuprous oxide taken at right angles to $1\bar{1}0$. All the points shown here by circles are allowed by the structure factor. If, however, we assume the surface of cuprous oxide to be of such a nature that one of the diffracted beams becomes a primary beam, additional spots in some cases would appear. In the case of the photograph shown in fig. 12 (Pl. III.), if it is assumed that a beam after diffraction from the (111) planes becomes a primary beam for a second diffraction, all the spots forbidden by structure factor can be accounted for. In fig. 13 the crosses show these spots due to the (111) as the primary beam. The same explanation applies to forbidden spots observed in fig. 11 (Pl. II.).

8. *Copper Sulphide on Copper.*

Since all sulphide solutions attack copper immediately they come in contact with it, the electrolytic circuit consisted of a copper crystal anode, cut to (111) face, hung in a porous pot full of slightly acidified distilled water surrounded in turn by a saturated solution of sodium sulphide contained in a beaker.

It was found that the thin films of sulphides show a single crystal orientation at a current density of about 0.09 ma./cm.² Increase in the time of treatment at this current density did not produce thicker films, showing that the deposit formed is slightly soluble in water.

Crystal Structure and Orientation Relationship.

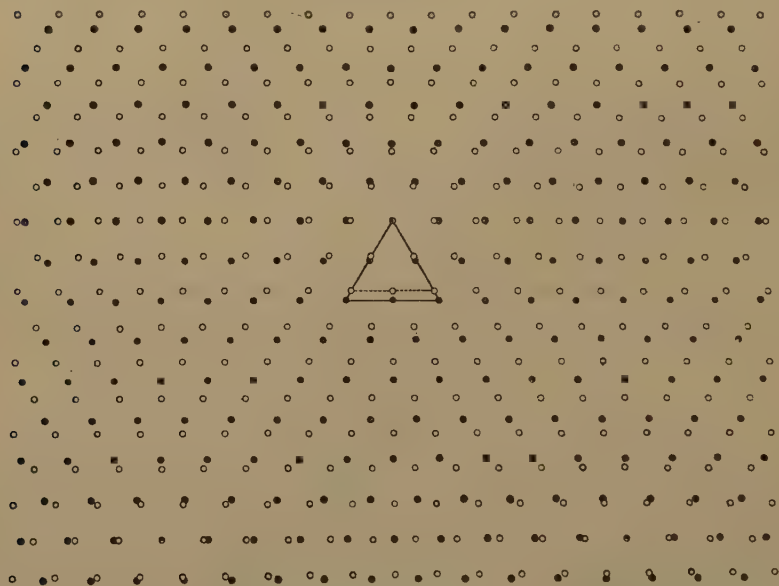
Two reflexion photographs from the film, when the copper crystal was set in the $1\bar{1}2$ and $1\bar{1}0$ azimuths, are shown in figs. 14 and 15 (Pl. III.) respectively. As to the orientation, it was observed that the spots due to the film followed the symmetry of the spots due to the copper base. On measurement, however, it was found that the spots of the film did not fit into any of the crystal structures of either cuprous or cupric sulphide, which, as worked out by X-rays, are orthorhombic ($a=11.8$ A., $b=27.2$ A., and $c=22.7$ A.) and hexagonal ($a=3.74$ A. and $c=16.1$ A.) respectively.

Attempts were therefore made to determine the type of the unit cell of the new sulphide with the help of electron diffraction patterns only. The two patterns in figs. 14 and 15 (Pl. III.) were found to repeat themselves at intervals of 60° , showing that the film had either a hexagonal or a cubic structure. Also, by regarding these patterns as two sections of the reciprocal lattice at right angles to one another, and by assuming that the reciprocal lattice is body-centred cubic, the new sulphide structure worked out to be cubic with an edge of 4.058 A. It also came out from the analysis that the (111) plane of the sulphide is parallel to the (111) plane of the copper substrate, with parallelism between the cube edges

of the two lattices. This orientation relationship is shown in fig. 16 with the positions of copper atoms in the growing sulphide lattice and the copper base*.

In connexion with the cubic structure proposed, it might be added that according to X-ray data on cubic crystals no substance is yet known to exist with an edge 4.058 Å. It is, therefore, likely that the substance in the film is some modification of copper sulphide. In fact the chemistry of such thin films has sometimes been found to be different from that of the same substance in bulk.

Fig. 16.



9. Copper Sulphide on Copper. *Dry Method.*

Experimental.

In order to further the study of the nature of the sulphide film, direct action of H_2S gas was tried on the etched surface of copper crystal cut to (111) face. The crystal specimen was put in an evacuated chamber, and dry low-pressure sulphuretted hydrogen passed slowly over it through a capillary tube. The resulting film was examined in the camera by the reflexion method.

* It may be mentioned that "spinel" twinning was observed in the pattern shown in fig. 15 (Pl. III.).

Even the thinnest films produced in this way did not show single crystal growth of the sulphide, but yielded, in all cases, ring patterns typical of polycrystalline formation. However, when a specimen treated in this way was put in a beaker containing a 10 per cent. solution of ammonium persulphate, the sulphide film was undermined and, after a few minutes, floated on the surface of the solution. It was picked up on a gauze and examined in the camera by transmission. This pattern in fig. 17 (Pl. III.) shows rings as well as spots or arcs superimposed on the rings. The spots show a definite symmetry of arrangement characteristic of single crystal growth.

TABLE II.

Cuprous Sulphide on Copper.

Camera length : 38.16 cm.

Wave-length : 0.0681 Å.

Action of H_2S gas on copper.

Cuprous sulphide rings.		
<i>r.</i>	<i>d.</i>	Intensity.
cm.	Å.	
0.790	3.290	Ft.
9.85	3.057	m.
1.080	2.406	Fuzzy.
1.320	1.968	St.
1.385	1.876	St.
1.501	1.731	m. (seems unresolved)
1.700	1.529	V. Ft.
1.890	1.375	V. Ft.

It seems that under favourable conditions of growth, when the chemical reaction is slow, only the lowest layers in contact with the surface of the copper crystal are oriented, so that in transmission a pattern both of spots and rings is obtained, the latter being due to polycrystalline layers at the top. Since such a film cannot be more than about 500 Å. thick, it appears that the copper base loses its influence on the growing sulphide film after it reaches that thickness.

Results.

The measurements of the ring pattern of fig. 17 (Pl. III.) are given in Table II. The interplanar spacings when worked out did not fit either the orthorhombic structure of cuprous sulphide or the hexagonal of the cupric sulphide as given by X-rays and mentioned before.

Although chemically, at ordinary room-temperatures, the reaction between sulphuretted hydrogen and copper must result in the formation of cuprous sulphide, the spots or the arcs in the transmission pattern show a marked hexagonal symmetry. The orthorhombic structure of the cuprous sulphide, therefore, does not seem to be a possibility. However, mention must be made here of similar arced transmission patterns which were obtained by Garrido and Hengstenberg (1932) when they studied films of paraffin $C_{21}H_{44}$. Although this paraffin has an orthorhombic structure, the hexagonal symmetry of the arcs was explained as due to the pseudo-hexagonal character of the paraffin lattice. However, the dimensions $a=11.8$ A., $b=27.2$ A., and $C=22.7$ A. of the orthorhombic cuprous sulphide lattice are not such as to make it possible to regard it as pseudo-hexagonal.

Now, if the structure is supposed to be hexagonal, the symmetry of the pattern suggests that the electron beam is parallel to the C-axis of the hexagon. Also the measurement shows that on this assumption the side of the basal plane of the unit hexagon is 1.136 A. or some integral multiple of it. This does not fit in with the hexagonal structure of cupric sulphide where the side of the hexagon is 3.74 A. Consequently, the structure of the sulphide film according to the study made here remains unknown.

Another interesting feature of the pattern shown in fig. 17 (Pl. III.) is the occurrence of two spots on two consecutive rings just on top of one another. Similar arrangement of spots in the transmission patterns of films of paraffin $C_{21}H_{44}$ was observed by Garrido and Hengstenberg (1932), who explained it as due to the twinning of the pseudo-hexagonal lattice of the paraffin. As has already been pointed out, such an explanation cannot apply to the lattice concerned here. It suggests, therefore, that either the two arcs lying on two different rings close to one another are two different orders of the same family of planes in the film, or they are due to two different planes inclined so that they may nearly be regarded as parallel.

The first suggestion is very unlikely because the dimensions of the cell have to be very large to give two consecutive orders of reflexion so close to one another. Moreover, the cell would have to have a very unusual structure factor to disallow the lower orders which do not appear on the plate. The second suggestion at present seems to be the only alternative. It is only to be assumed that the hexagonal cell of the film, although it has a small base of edge 1.136 A. (or some multiple of it), has considerable length along the C-axis, to make some planes which cut the C-axis nearly parallel to the hexagonal planes. Since no spot pattern was ever obtained by the reflexion method, it was not possible to determine the length of the C-axis of the unit hexagon of the sulphide lattice.

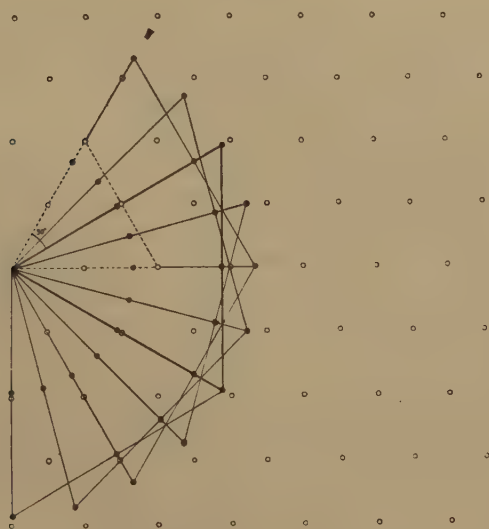
10. Discussion.

The experiments indicate that the deposits formed electrochemically on a copper crystal anode are capable of growing as single crystal lattices provided the current density is low and the reactions are slow.

In almost all the cases, when the deposit is formed, the atoms of copper from the substrate migrate to new positions in the deposited lattice and form an expanded face-centred cubic framework.

A close examination of figs. 7, 10, and 16 reveals that, although the atoms do not fit exactly, single crystal orientation of the growing lattice is visible in the diffraction patterns. It is found that Rüdiger and

Fig. 18.



Brück's law, which holds good for only a few first cells, is valid in these cases. Moreover, the fact that two orientations for the same compound were never observed in the patterns, and also the fact that, in spite of similar cubic structure, the halide films show one type of orientation (figs. 7 and 10) and those of oxide and sulphide (porous pot) another type (fig. 16), further confirms the validity of Brück's law in these cases.

In order to see if there is a theoretical arrangement of atoms which would give a better fit than that given by Brück's law, fig. 18 was drawn showing the positions of the copper atoms in the (111) plane of cuprous iodide and those in the (111) plane of copper, when the former was twisted in its own plane through angles other than 30° (the observed orientation) with respect to the (111) plane of the copper base. No angle

other than 30° gave a better fit of atoms in the two planes of the deposit and the substrate. Hence Brück's law would explain the orientations observed in our patterns.

However, this law can account only for the growth of a few cells and cannot hold in general. Since the periodicities are incommensurable, the atoms which showed a close fit at one point would get out of step and shift to unfavourable positions as the crystal grows. This can be seen in all the three figures (7, 10, and 16). Unless an exhaustive study is made of the growth of one lattice on another having completely incommensurable dimensions, and more data on orientation relationship available, no definite conclusions can be drawn as to the mechanism of the growth of the deposit on the substrate.

11. *Summary.*

Thin films of cuprous iodide, bromide, sulphide, and oxide have been formed electrolytically on an etched surface of a single crystal of copper. It is shown, by diffracting electrons from the surfaces of these films, that they grow as single crystals when deposited at small current densities, revealing a definite orientation relationship to the copper crystal base. All of them have their most densely packed (111) planes parallel to the (111) plane of copper. The cuprous oxide film shows spots forbidden by the structure factor, which are attributed to a second diffraction of the primary beam. Cuprous iodide, bromide, and sulphide films show additional spots in the $1\bar{1}0$ azimuth due to spinel twinning, (111) being the twin plane.

Copper sulphide film formed electrolytically has a structure which does not agree with the structure of either cuprous or cupric sulphide as given by X-rays. Instead, it shows cubic symmetry with a unit cell of edge $4.058 \text{ \AA}$.

Copper sulphide films formed by the direct action of hydrogen sulphide gas on copper crystal show no single crystal formation in reflexion patterns; transmission of electrons through such films when stripped, however, show a spot pattern with hexagonal symmetry—again having a structure different from that given by X-rays. The two sulphides, one formed electrolytically and the other by dry hydrogen sulphide gas, had two different structures.

12. *Acknowledgment.*

In the end I should like to thank Professor G. P. Thomson for his guidance and valuable suggestions throughout this research.

References.

- Thomson, G. P., Proc. Roy. Soc. A, cxxxiii. p. 1 (1931).
Thomson, G. P. and Fraser, Proc. Roy. Soc. A, cxxviii. p. 640 (1930).

- Cochrane, *Proc. Phys. Soc.* xlviii. p. 723 (1936).
Lassen, *Phys. Zeit.* xxxv. p. 172 (1934).
Brück, *Ann. d. Phys.* xxvi. p. 233 (1936).
Lassen and Bruck, *Ann. d. Phys.* xxii. p. 65 (1935).
Rüdiger, *Ann. d. Phys.* xxx. p. 505 (1937).
Yamaguti, *Proc. Phys. Math. Soc. Jap.* xx. p. 230 (1938).
Miyake, *Sci. Pap. Inst. Phys. Chem. Res. Tokyo*, xxxiv. p. 565 (1938).
Ewald, *Zeit. Krist.* lvi. p. 129 (1921).
Raether, *Z. Phys.* lxxviii. p. 527 (1932).
Garrido and Hengstenberg, *Zeit. für Krist.* lxxxii. p. 477 (1932).

IX. *On the Presentation of Crystal Chemistry.*

By A. F. WELLS, M.A., Ph.D.

(Crystallographic Laboratory, Cambridge) †.

[Received May 1, 1940.]

IN a crystal atoms are arranged in a more or less regular manner, forming a solid body with characteristic physical and chemical properties. The actual arrangement of the atoms relative to one another results from the forces acting between the atoms, and these forces must be interpreted in terms of the electronic structures of the constituent atoms. It is logical, therefore, first to consider atomic structure and the types of bonds formed between atoms. We then have to consider the interrelations of the various structure types and to relate the properties of the crystal to its structure. In order to present the subject-matter of crystal chemistry in an intelligible and orderly manner, it is necessary to have some sort of a classification of crystal structures. In determining the structure of a crystal we assume that atoms of known electronic configuration exist in it, and we find their positions. One possible kind of classification of structures is therefore a geometrical one, in which the emphasis is placed on the relative positions of the atoms. From the interatomic distances and the disposition of the atoms relative to one another it is often possible to infer the nature of the bonds operating between given atoms, and independent evidence concerning bond type comes from physical and chemical properties and from purely theoretical studies. Accordingly, it should be possible to classify crystal structures according to the nature of the bonds within the crystal. Classification according to empirical formulæ is mentioned below.

In considering the merits of the various methods of classification it is important to bear in mind the object of the classification, which in the present case is to provide the basis for the presentation of crystal chemistry. Other types of classification may be required for different purposes, as, for example, in the "Strukturbericht."

By a purely geometrical classification we mean one in which we are interested only in the co-ordinates of the atoms. For reasons of space it is not possible even to mention here all the previous work in this field; a good bibliography is given by Nowacki (1935). One feature of the

† Communicated by the Author.

classification proposed in the present paper is the broad division of structures into four classes, according to the extent of the complexes which may be distinguished in the structures. This division was made by Weissenberg, and the concept has been elaborated by Niggli and others. The complexes themselves—finite, one-dimensional (chains), two-dimensional (layers) and three-dimensional—may be derived and classified in two essentially different ways. They may be derived from the equivalent positions of space-groups (see, for example, Hermann, 1928–9), or they may be derived without reference to point symmetry (Niggli, 1933, Wells, 1940). The detailed development of the geometrical possibilities in the various classes of complexes will not be considered in the following account. Any comprehensive classification of crystal structures must allow for defect lattices which can be satisfactorily classified only according to some geometrical scheme (see Laves, 1930 and 1939, Ströck, 1936, Nowacki, 1938).

According to classical structure theory the atoms in a crystal are arranged at definite points in the lattice consistent with the symmetry of one of the 230 space-groups. It follows that all unit cells of the structure should contain the same numbers and kinds of atoms arranged in the same way. The theory also requires that a given set of equivalent positions must be occupied, and completely occupied, by atoms of the same sort. Structures may depart from these requirements in one or both of the following ways :—

- (1) The mean † position of a given atom (or atoms) cannot be located at a definite point.
 - (a) There is random motion of atoms (translation).
 - (b) The locus of the centre of the atom is stationary, *e. g.*, circle or surface of sphere (atom off the axis of rotation of a rotating group).
- (2) The mean position of every atom is a fixed point, but the content of a given unit cell is not the same as that of every other, differing in number, position and/or kind of atom. This includes all structures containing atoms arranged at random. Two main subdivisions may be made :—
 - (a) Each set of equivalent positions, if occupied at all, is completely filled.
 - (b) Some of the sets of equivalent positions are only partially occupied.

† Vibration of an atom about a fixed mean position (temperature motion) does not constitute a departure from structure theory.

These two classes, 2 *a* and 2 *b*, the "complete" and "incomplete" lattices of Laves (1930), are further subdivided by Strock (1936). We shall not include here the detailed crystallographic analysis of defect structures, since a rigid treatment of this problem involves classifying all solid bodies and the resulting classification is unlikely to be of value as a basis for the presentation of crystal chemistry (compare, for example, that of Nowacki, 1938). Instead, we shall indicate some of the possible types of defect structure when dealing with the normal structures to which they are most closely related.

We now come to schemes in which the primary divisions are made according to bond type. Four types of interatomic binding have long been recognized, ionic, homopolar, metallic, and van der Waals, but a division of structures into the corresponding four classes is inadequate, since there are few structures in which all the interatomic forces are of the same kind, and fewer still in which the bonds are purely ionic, homopolar, etc., in character. These complications are discussed in Appendix I., where it is shown that classifications of this sort become compromises between geometrical and bond type classifications, and in spite of their complexity are still not comprehensive.

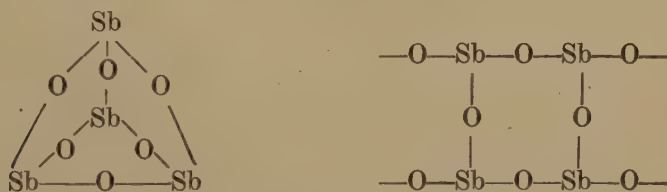
Treatment according to empirical formulæ is valuable when considering the various structure types possible for simple binary and ternary inorganic compounds and the conditions governing the transition from one structure type to another (see p. 125 *et seq.*). As a general method of classification it has little value except for cataloguing purposes.

Proposed Classification of Crystal Structures.

The following scheme is the result of an attempt to provide a classification of crystal structures, avoiding some of the inconsistencies and complications of existing classifications (see Appendices). Although it will group together in many of its subsections the same types of structure as do other schemes, it is designed to allow a discussion of the actual nature of the bonds in a given structure without having prejudged the issue. It should be borne in mind that the adoption of a classification of this sort does not imply that compounds must be described in isolated groups determined by their geometrical characteristics. The treatment of simple AX_n structures suggested below (p. 127) indicates the way in which the various structural types may be related to one another. Our first division is as follows :—

1. Crystals containing finite complexes.
2. " " infinite one-dimensional complexes.
3. " " " two- " "
4. " " " three- " "

The assignment of a crystal to one of these classes is generally possible from a consideration of the geometry of the structure, supported by the properties of the crystal. The recognition of complexes rests largely on interatomic distances, since the forces between the atoms of one complex and those of another are in general ionic, van der Waals or hydrogen† bonds (see, however, references to the structures of As, Se, Sb_2S_3 , etc., below). The anionic and van der Waals radii of an atom are similar in magnitude and are much greater than the covalent radius. The H bond gives rise to an intermediate interatomic distance, and the fact that this type of bond was recognized in crystals in this way justifies its inclusion in our general scheme, a characteristic of which is the fact that structures are broadly classified on the basis of experimentally determined distances. The usual classification of silicates, according to the extent of the grouping Si_xO_y , is incorporated, and the relation between the polymorphic forms of an element or compound is emphasized. For example, elements forming two bonds may form chains or rings, as in the cases of S and Se. The two simplest configurations for Sb_2O_3 , in which Sb forms three and O two bonds, are the tetrahedral molecule Sb_4O_6 and the infinite chain $(\text{Sb}_2\text{O}_3)_n$, shown in idealized form below:—



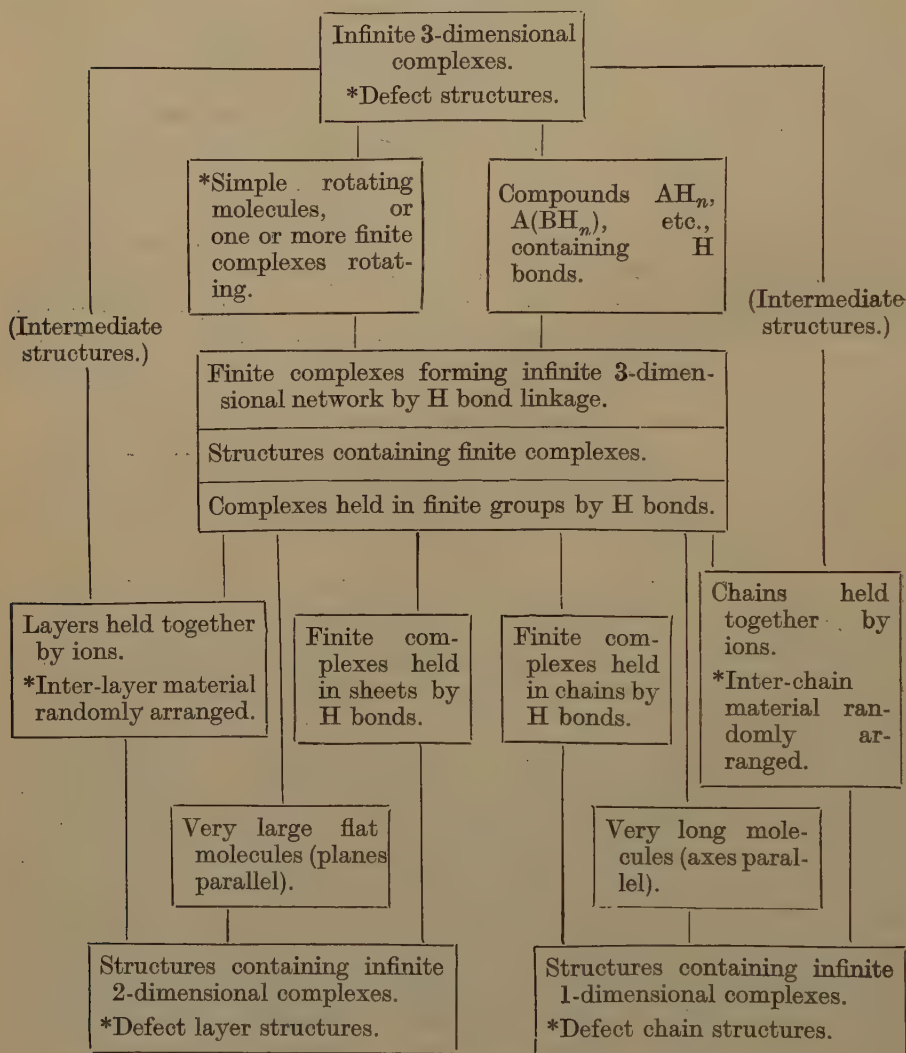
The former occurs in the cubic variety, senarmonite, and the latter in the rhombic form, valentinite. This relation between the tetrahedron and the double chain may be compared with that between the ring and the simple chain.

The presentation of crystal chemistry implies far more than an orderly description of crystal structures, and it is realized that the inter-relationships and analogies between the various structure types cannot all be expressed on a single two-dimensional chart. It is clearly of little value merely to point out that a certain compound crystallizes with a particular structure, *e. g.*, SiS_2 with a chain structure. Some attempt must be made to account for the adoption of a particular structure type which will be compared with those of similar compounds (*e. g.*, SiO_2 , GeS_2). Furthermore, owing to the fact that in the proposed classification chemically

† When used in this general way the term "hydrogen bond" includes the "hydroxyl bond" (see below).

related compounds like borates and silicates are scattered among the classes, short summarizing reviews of the crystal chemistry of important elements are very valuable. The general scheme is set out in the accompanying chart, and the various classes will now be considered in the following order :—

CHART.



N.B.—The groups of defect structures are distinguished by asterisks, as in the text.

1. Crystals containing *finite* complexes.

I. Structural units all of the same kind.

- (a) van der Waals bonds between complexes.
- (b) Hydrogen bonds between complexes leading to :
 - (i.) Finite groups.
 - (ii.) Infinite chains.
 - (iii.) „ layers.
 - (iv.) „ 3-dimensional network.

II. Structural units of more than one kind.

- (a) van der Waals bonds between complexes.
- (b) Hydrogen bonds „ „

Hydrates.

(1) Metallic Salts.

- (i.) Without H bonds.
- (ii.) With H bonds.

(2) Non-ionic compounds.

- (i.) Without H bonds.
- (ii.) With H bonds.

(3) Hydrates which can be reversibly hydrated.

- (c) Ionic bonds between complexes.

2. Crystals containing *infinite 1-dimensional* complexes.

I. Infinite 1-dimensional complexes only.

- (a) van der Waals bonds between chains.
- (b) Hydrogen „ „ „
- (c) Ionic „ „ „

II. Infinite 1-dimensional complexes and finite complexes or ions.

3. Crystals containing *infinite 2-dimensional* complexes.

I. Infinite 2-dimensional complexes only.

- (a) van der Waals bonds between layers.
- (b) Hydrogen „ „ „
- (c) Ionic „ „ „

II. Infinite 2-dimensional complexes and finite complexes or ions.

4. Crystals containing *infinite 3-dimensional* complexes.

- (a) Inert gases.
- (b) Metals and alloys.
- (c) Ionic and homopolar crystals.

1. CRYSTALS CONTAINING FINITE COMPLEXES.

We may mention here an important difference between crystals in this class and those with infinite complexes. Infinite complexes do not retain their individuality in other states of aggregation, whereas finite complexes which exist in crystals may be stable in solution or in the vapour state. Their stability will of course depend on the type and

strength of the bonds within the complexes, which in turn are influenced by temperature, interaction with the solvent and so on.

I. *Structural Units all of the Same Kind.*

According to the nature of the packing and the distances between atoms of different complexes we may distinguish two sub-groups.

(a) Crystals in which only van der Waals forces operate between complexes.

These structures are determined by the most efficient packing of units of a given shape held together by undirected forces (*e. g.*, naphthalene). When the complex is roughly spherical the same structure types may occur as in crystals with three-dimensional complexes, a group of atoms replacing the single atom—compare the structures of Sb_4O_6 and diamond. As the shape of the complex deviates from spherical so the structures bear more analogy to those of crystals containing infinite one- and two-dimensional complexes. An extreme example is the hydrocarbon $\text{C}_{1000}\text{H}_{2002}$, which for most purposes may be regarded as an infinite chain. (The X-ray powder patterns of long chain hydrocarbons remain practically constant for molecules over 130 atoms long—Bunn, 1939.)

(a*) Structures in which the finite groups are rotating constitute an important class of defect structures. If the group becomes effectively spherical as the result of rotation, then highly symmetrical close-packed structures result (*e. g.*, H_2). In the long chain hydrocarbons rotation about the chain axis begins below the melting-point, and a comparison of the hexagonal high temperature form of $\text{C}_{29}\text{H}_{60}$, for example, with the low temperature orthorhombic form (in which there is no rotation of the chains) illustrates the change in packing due to rotation of the molecules.

(b) Crystals in which hydrogen bonds operate.

Since the number of hydrogen bonds is limited, the forces between many of the atoms of different complexes are, of course, van der Waals bonds. These crystals are characterized by shorter distances between certain pairs of atoms belonging to different complexes than between other pairs. These directed inter-complex forces are stronger than the van der Waals bonds, as shown by gentle treatment of the crystal, which breaks the latter but leaves intact the hydrogen bonds (*e. g.*, basal cleavage of pentaerythritol), but weaker than the forces holding together the atoms of the complex itself (*e. g.*, dissolution of a crystal of pentaerythritol breaks the hydrogen bonds without disrupting the molecule).

* The groups of defect structures are distinguished by asterisks.

The atoms between which H bonds operate act, so far as packing is concerned, like localized charges, leading to specific structure types (resorcinol, pentaerythritol) rather than the most economical packing characteristic of non-polar molecules like hydrocarbons. The analogy in three-dimensional complexes is found in the open structure of diamond, resulting from the tetrahedrally directed carbon bonds, compared with the close packed structures of inert gases and true metals. The presence of H bonds leads to aggregates of finite complexes which may be :—

(i.) Finite.—The association of carboxylic acid molecules in pairs has not yet been demonstrated for a crystalline acid, but is well established for many acids in the dissolved state and for formic acid in the vapour (Pauling and Brockway, 1934).

(ii.) Infinite chains, in diketopiperazine, β -oxalic acid.

(iii.) Infinite layers, in boric acid, α -oxalic acid, pentaerythritol.—The cleavages of the two forms of oxalic acid, into laths and flakes, illustrate the effect of H bonds on physical properties.

(iv.) Infinite three-dimensional network.—The simplest example is the structure of ice, in which all the bonds are H bonds. The structures of the two forms of $\text{Te}(\text{OH})_6$ provide further examples.

The resultant groupings in the first three classes are then held by residual forces. The structures in (ii.) and (iii.) are similar in many ways to chain and layer structures respectively, and this is indicated by the positions of these sub-groups on the chart. The introduction of this second kind of inter-complex binding increases the number of possible structure types in the way shown in fig. 1. Atoms A may form finite complexes A_n (A_2 is taken as an example), infinite chains or layers. Only the simplest cases are shown and three-dimensional complexes are entirely omitted from the figure. The finite complexes may then associate as in fig. 1 (b).

More complex cases may arise according to the shape of the molecule and the position of the polar groups. Thus, long chain carboxylic acids associate in pairs and the pairs then stack in layers with their axes parallel. Residual forces hold together the molecules in the layers and are also responsible for the cohesion of the layers.

II. Structural Units of more than One Kind.

As before, we may subdivide these structures into groups corresponding to the type of inter-complex binding: (a) van der Waals, (b) hydrogen, or (c) ionic bonds.

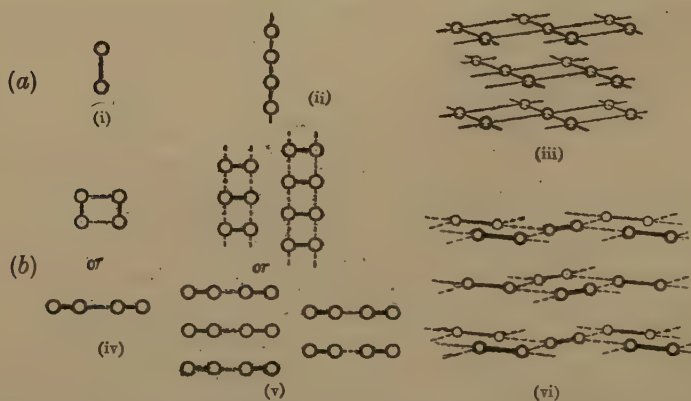
(a) van der Waals bonds between complexes.

No complete structure determination of a crystal of this group appears to have been made, but molecular lattices such as $\text{CHI}_3 \cdot 3\text{S}_8$ are presumably of this type.

(b) Hydrogen bonds between complexes.

It is convenient to divide these structures into two groups according to whether all the complexes are joined by H bonds or only those of one sort. This division brings out the geometrical analogy between the structure of, for example, $\text{Te}(\text{OH})_6$ on the one hand, and those of $(\text{NH}_4)_2\text{H}_3\text{IO}_6$ and KH_2PO_4 on the other. In crystals of $(\text{NH}_4)_2\text{H}_3\text{IO}_6$ H bonds link the IO_6 octahedra into a three-dimensional framework with NH_4 ions occupying the interstices. In KH_2PO_4 there is a similar

Fig. 1.



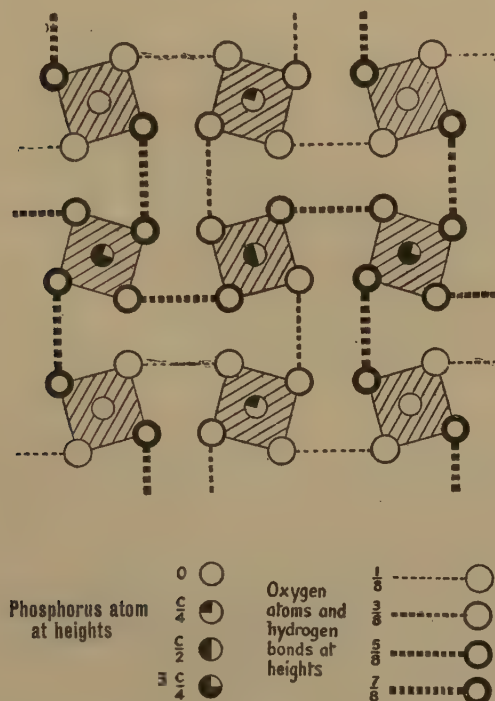
In (a) are shown simple examples of (i.) finite, (ii.) infinite one-dimensional, and (iii.) infinite two-dimensional complexes. In the latter three layers are shown in perspective. Hydrogen bonds (broken lines) may then link the finite complex (i.) into (iv.) finite, (v.) infinite one-dimensional, or (vi.) infinite two-dimensional aggregates, of which (vi.) is shown in perspective.

linking of PO_4 tetrahedra by H bonds with K ions in the holes between eight oxygen atoms. The O : H ratio in each of these compounds is 2 : 1, and their stability is due to the formation of the H_3IO_6 and H_2PO_4 frameworks. The latter is illustrated in projection in fig. 2, in which the H bonds are approximately parallel to the plane of the paper (001).

Sodium bicarbonate is an example of a structure in which the anions are joined into chains by H bonds, and from the geometrical point of view may be compared with diopside. Similar divisions may therefore be made as in I. (b), (i.)–(iv.).

In contrast to the above cases there are certain ammonium salts in which the H bonds operate between the anions and the cations, viz., NH_4F , NH_4N_3 and NH_4HF_2 . A comparison of the structure of the latter with that of KHF_2 shows the striking alteration in structure caused by H bond formation.

Fig. 2.



Part of the structure of KH_2PO_4 projected on the basal plane. The broken lines represent hydrogen bonds which link the PO_4 tetrahedra into an infinite three-dimensional network.

Hydrates.

The behaviour of water as a molecule with a tetrahedral charge distribution and its ability to take part in H bond formation make it desirable to deal separately with hydrates rather than to treat the molecule H_2O as merely an additional complex in the structure. It may also be pointed out that many salt hydrates may be satisfactorily represented by an electrostatic picture.

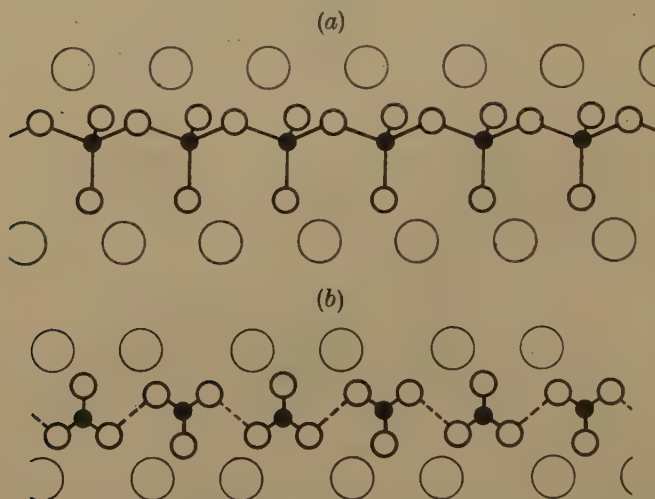
(1) Hydrates of metallic salts.

In these structures some or all of the water of crystallization is associated with the metal atoms forming well-defined co-ordination groups :

(i.) Without H bonds.

Where the anion is not sufficiently electronegative or its electro-negative nature is lost through incorporation in a complex group H bonds are not formed, as in the structures of $\text{NiSnCl}_6 \cdot 6\text{H}_2\text{O}$ and isomorphous salts.

Fig. 3.



Projections of portions of the structures of (a) diopside, $(\text{Ca}, \text{Mg})\text{SiO}_3$, and (b) sodium bicarbonate, NaHCO_3 . Large circles represent metal atoms, small open circles oxygen atoms, small black circles (a) silicon and (b) carbon atoms. Broken lines indicate hydrogen bonds.

(ii.) With H bonds.

In certain halides and oxy-salts H bond formation is possible between the halogen or oxygen atoms of the oxyanion and the oxygen atoms of the water molecules. The structures are very varied and sometimes of considerable complexity, and might be classified according to the functions of the water molecules. For example, in both the hexa- and hepta-hydrates of NiSO_4 six water molecules form a co-ordination group around the nickel atom, but in the latter compound the seventh water molecule is bound to three other water molecules and to one oxygen atom of a sulphate ion. In gypsum the Ca and SO_4 ions form well-defined layers, the water molecules lying between them and holding them together.

(2) Hydrates of non-ionic compounds.

These may be similarly subdivided according to whether or not H bonds are formed between the molecules of the compound and the water molecules.

- (i.) In the unstable hydrates of the inert gases, halogen derivatives of paraffins and hydrocarbons, there is very weak cohesion between these molecules and the water molecules, as shown by the high dissociation pressures of the hydrates.
- (ii.) In contrast to these, the hydrates of carboxylic acids, for example, have considerable stability owing to the fact that the water molecules form H bonds with the oxygen atoms of the carboxyl groups (as in $(\text{COOH})_2 \cdot 2\text{H}_2\text{O}$).

A characteristic feature of the hydrates so far described is that removal of the water destroys the crystal, and the structure of the anhydrous compound (or element), if it exists, bears no relation to that of the hydrate. For completeness we may add here a third group of structures in which the water molecules are loosely held and from which they may be removed without causing any radical alteration in the structure.

(3) Structures which may be reversibly hydrated and dehydrated.

The effect on the structure of dehydration depends on the geometrical nature of the structure. The rigid silicon-oxygen framework of zeolites, for example, is unaffected by the removal of water (or other included material), whereas layer structures such as clay minerals may contract appreciably in one direction owing to the closer approach of the layers.

(c) Ionic bonds between complexes.

This class includes all salts of organic and inorganic acids containing finite complex ions and ionic co-ordination compounds generally. It is well to consider here the definition of "complex ion." Organic ions are easily recognized on account of the strong homopolar bonds between the atoms of the complex. For inorganic ions we may adopt the simple physical definition of Bernal (1934): "a charged group of atoms (closed or open) for each atom of which the forces attaching it to the group are greater than the forces attaching it to other atoms." It should be noticed that we specifically omit any mention of bond type. The classification of ionic structures according to electrostatic bond strength, which has been developed by Evans, is reviewed in Appendix II.

Two points may be noted here. Firstly, structure types may occur

similar to those of AX_n compounds with three-dimensional complexes, for example :—

CaC_2 —NaCl structure	$[Mg(NH_3)_6]I_2$ — CaF_2 structure
calcite—NaCl ,,	$[Al(NH_3)_6]Cl_3$ — YF_3 ,,
aragonite—NiAs ,,	

Where the complexes are approximately spherical in shape the structure may be strictly similar to that of the analogous AX_n compound, in which case radius ratio criteria are applicable (see the discussion of radius ratio in 4 c). If the complex group is appreciably anisotropic in shape the structure is a distorted version of the simple AX_n structure, as in the cases of CaC_2 and $CaCO_3$. Secondly the stability of a complex group depends on its environment and on the physical conditions. As examples we may quote the crystallization of Li_2MoO_4 at low temperatures with the phenacite structure and of Ag_2MoO_4 with the spinel structure, and the change in structure of Li_2MoO_4 , from phenacite to spinel, with rise of temperature. These examples show the close relation between crystals containing finite and infinite three-dimensional complexes and at the same time illustrate one of the weaknesses of a classification based on electrostatic bond strength (see Appendix II.).

(c*). As in I. (a*) we differentiate structures in which one or more complexes are rotating from those in which all the atoms can be assigned fixed mean positions. In the case of all molecules, except apparently H_2 , rotation ceases at a sufficiently low temperature. It may be objected that it is undesirable to place the same compound in different classes according to the temperature, particularly since at all temperatures above the absolute zero atoms in crystals possess some vibrational energy, and rotation sets in only when the potential energy exceeds a certain critical value. It is, however, worth while to make the distinction because in the one case it is possible to locate the mean position of each atom at a definite point in the structure. On the other hand, the oxygen atoms of a NO_3 ion rotating about its triad axis describe circular orbits, and structures containing such ions would be classed as defect structures.

2. CRYSTALS CONTAINING INFINITE ONE-DIMENSIONAL COMPLEXES.

Here we are concerned with crystals in which chains or bands of atoms extend throughout the crystal, the chains being recognizable on account of the larger distances between atoms of different chains than between the nearest atoms within the chains. This, in general, implies that the forces between the chains are different in nature from those acting between adjacent atoms in a chain. It is usual to describe ionic crystals, particularly those with complex structures, in terms of the arrangement of the

coordination polyhedra (of anions around the cations) and the way in which these polyhedra share edges and corners. For example, in pseudobrookite, Fe_2TiO_5 , we may distinguish chains with the composition TiO_5 formed of TiO_6 octahedra sharing two opposite corners. The iron atoms between the chains are, however, also surrounded octahedrally by six oxygen atoms, and since there is no reason to suppose that the $\text{Ti}-\text{O}$ and $\text{Fe}-\text{O}$ bonds differ appreciably in type, chains of this kind are not recognized as infinite one-dimensional complexes in this classification. As in the case of structures containing finite complexes, chain structures will be subdivided according to the nature of the inter-complex forces. In other words, we emphasize the relation of the chains, as units, to one another rather than their detailed internal structure. The latter must be discussed in each case in relation to the properties of the constituent atoms. The derivation of the various possible arrangements of atoms in infinite chains is a separate problem which will not be considered here.

I. Structures containing only infinite one-dimensional complexes fall naturally into a number of groups, as indicated on the chart.

(a) Van der Waals bonds between chains, as in SiS_2 , PdCl_2 , SeO_2 , Sb_2O_3 , etc. There are no known examples containing chains of different compositions. The structures of selenium and tellurium and of Sb_2S_3 and Bi_2S_3 deserve special mention owing to the semi-metallic properties of the crystals. In the two former structures the chains are well defined, as is seen from the distances between an atom and its nearest neighbours in the same chain and its next nearest neighbours in adjacent chains. The published interatomic distances in Sb_2S_3 , particularly the large variation in $\text{Sb}-\text{S}$ from 2.38–2.67 Å., are, however, less satisfactory. Just as the structures of arsenic, antimony and bismuth show a gradual transition from the two- to the three-dimensional complex, so we may expect to find analogous intermediate cases between one- and three-dimensional complexes, and allowance is made for these on the chart.

(b) Hydrogen bonds between complexes. Proteins consisting of infinite polypeptide chains are probably of this type.

(c) A further possible group would contain chain ions with opposite charges.

II. Structures containing one-dimensional complexes and finite complexes or ions.

A number of ionic structures fall in this category, including the fibrous silicates (pyroxenes and amphiboles), $\text{Ca}(\text{B}_2\text{O}_4)$, and $\text{NH}_4(\text{CdCl}_3)$.

I*. Corresponding to the above classes there may be defect structures of various kinds, though no examples are yet known; for example, structures containing rotating chains or chains randomly arranged. Other kinds of defect structures may be visualized corresponding to mixed crystals of isomorphous compounds with three-dimensional complexes, *e. g.*, (Na, K)Cl. For example, PdCl_2 forms a chain structure. Since Pt and Pd have the same covalent radius, we may imagine mixed crystals in which (a) all the metal atoms in a given chain are either Pt or Pd but the PtCl_2 and PdCl_2 chains are randomly arranged, or (b) each chain contains both types of metal atom distributed at random.

3. CRYSTALS CONTAINING INFINITE TWO-DIMENSIONAL COMPLEXES.

Similar divisions may be made to those in the previous class.

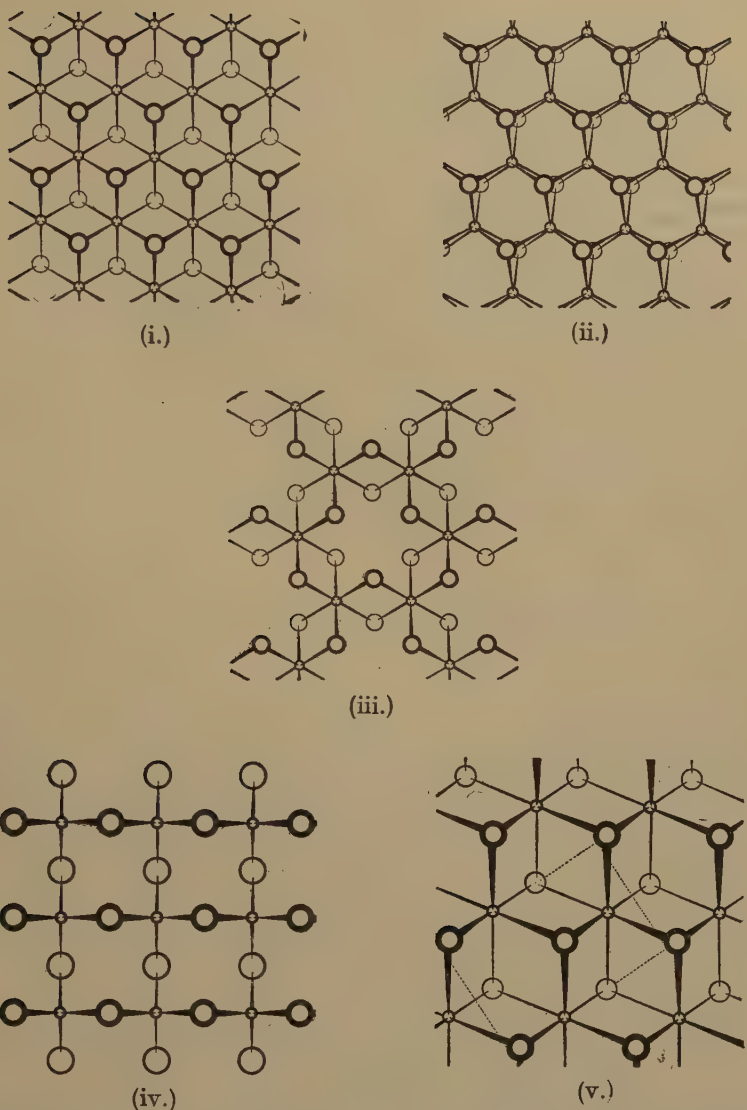
I. Structures containing two-dimensional complexes only include the following :—

(a) Van der Waals bonds between layers.

Examples : Many chlorides, bromides, and iodides AX_2 and AX_3 , and a number of sulphides, oxyhalides and hydroxides. In the corresponding oxides and fluorides, in which the A—X bonds are more ionic, three-dimensional complexes are more usual. The structures of some of the simple AX_2 and AX_3 layers are illustrated in fig. 4.

- (i.) The layers in the CdCl_2 and CdI_2 structures are similar, the structures differing in the way in which the layers pack together. Other compounds with the CdCl_2 structure include the dichlorides of Fe, Co, and Ni and NiI_2 ; with the CdI_2 structure, the diiodides of Ca, Mg, Mn, Fe and Co, the dibromides of Mg, Fe and Co, the disulphides of Sn, Ti and Zr, and the dihydroxides of Ca, Mg, Cd, Mn, Fe, Co and Ni. With reference to the latter see Bernal and Megaw (1935). For CdBr_2 and NiBr_2 see below.
- (ii.) The characteristic of the structure of MoS_2 (and of the isomorphous WS_2) is the disposition of the Mo bonds towards the corners of a trigonal prism (contrast the octahedral coordination in CdCl_2 and CdI_2).
- (iii.) The structure of the layers in the CrCl_3 and BiI_3 structures is closely related to that of the CdCl_2 and CdI_2 layers, as shown in the figure. The following halides also crystallize with the BiI_3 structure : SbI_3 , AsI_3 , FeCl_3 and CrBr_3 .
- (iv.) The HgI_2 layer is the analogue of (i.) for an element forming tetrahedral bonds.

Fig. 4.



The structures of some simple AX_2 and AX_3 layers. In each diagram the small circles represent metal atoms in the plane of the paper and the larger circles halogen, oxygen or sulphur atoms which lie in planes above and below that of the metal atoms. The layers shown are (i.) $CdCl_2$ (and CdI_2), (ii.) MoS_2 , (iii.) $CrCl_3$, (iv.) HgI_2 and (v.) $HgBr_2$. The non-metal atoms are projected normally on to the plane of the metal atoms except in (ii.), where the sulphur atoms are slightly displaced to avoid superposition.

(v.) The HgBr_2 layer is closely related to (i.) and is shown in the figure in a similar orientation. The structure may be regarded as intermediate between a layer and a molecular lattice, since two of the six bromine neighbours of a mercury atom are much closer (2.48 Å.) than the remaining four (at 3.23 Å.).

(b) Hydrogen bonds between layers.

The structures of a number of hydroxides differ from those in (a) in two ways; firstly in the way in which the layers are packed together, and secondly in the distance between OH groups of adjacent layers. The arrangement of the OH groups in a layer of $\text{Mg}(\text{OH})_2$ and of $\text{Al}(\text{OH})_3$ is similar, but whereas in the former structure the OH groups of neighbouring layers are close packed, in the latter a much more open packing is found. Furthermore, the distance between OH groups in adjacent layers in $\text{Mg}(\text{OH})_2$ is 3.22 Å. and in $\text{Al}(\text{OH})_3$ only 2.80 Å. These facts indicate the presence of directed O—H—O bonds between the layers, the "hydroxyl" bonds of Bernal and Megaw (1935). For a similar example, compare the structures of lepidocrocite ($\gamma\text{-FeO.OH}$) and FeOCl .

(c) Ionic bonds between layers.

No representative of this class, containing infinite two-dimensional ions of opposite charge, is known.

Intermediate structures.

The structures of Zn and Cd and of As, Sb and Bi may be regarded as intermediate between those containing infinite two- and three-dimensional complexes. In zinc and cadmium each atom has six nearest neighbours in the basal plane and six further removed, three above and three below the plane of the others. The respective distances of these two sets of neighbours are 2.660 and 2.907 Å. in zinc and 2.973 and 3.287 Å. in cadmium, indicating stronger bonds in, than between, the layers as suggested by the physical properties. In arsenic ("metallic" form) there are fairly well-defined layers, each atom having three neighbours in the layer at a distance of 2.51 Å., and three in an adjacent layer at 3.15 Å. In Sb and Bi the layers become progressively less well defined, as shown by the corresponding pairs of distances (2.87 and 3.37 Å. in Sb, 3.10 and 3.47 Å. in Bi), the structure tending to that of a close packed metal.

II. Structures containing infinite two-dimensional complexes and finite complexes or ions.

Of structures containing infinite layer ions the micas are the only large class, the complex layers being interleaved with metal ions. Some of the

related clay minerals may be of this type: others are probably defect structures. Certain graphitic compounds are presumably of this class, such as anhydrous graphitic acids and compounds with the alkali metals. For example, compounds of the type C_8K ($a=4.94$, $c=21.34$ Å.) and $C_{16}K$ ($a=4.94$, $c=17.45$ Å.) consist of graphite layers interleaved with various numbers of alkali metal atoms. The hydrated graphitic acids, on the other hand, show continuously varying c spacings according to water content, and are preferably described under II*.

I*. Defect structures in which the layers are rotated or translated with respect to one another. Examples are $CdBr_2$ and $NiBr_2$ recrystallized from alcohol (the sublimed material has the normal $CdCl_2$ structure).

II*. Defect structures in which the inter-layer material is randomly arranged. (See remarks above.)

4. CRYSTALS CONTAINING INFINITE THREE-DIMENSIONAL COMPLEXES.

This class comprises the following groups:—

- (a) Inert gases.
- (b) Metals and alloys.
- (c) Ionic and homopolar crystals, including the “framework” structures.

(a) The first group calls for little discussion. A number of crystals containing rotating diatomic molecules are structurally similar, for example the close-packed forms of N_2 , CO , HCl , and it is immaterial whether they are placed in this group or with the crystals containing finite complexes. When rotation is absent (HI , halogens) less close-packed structures result.

(b) The complexity of metal systems makes any rigid classification of doubtful value, but the broad division of metals into two classes:

- A sub-groups and transition metals (“true” metals),
- B sub-groups,

is on the whole consistent with their structures and properties. Inter-metallic compounds are then considered in three groups:—

- Alloys of $\left\{ \begin{array}{l} \text{(i.) Two true metals (AA).} \\ \text{(ii.) True metal with a B sub-group metal (AB).} \\ \text{(iii.) Two B sub-groups metals (BB).} \end{array} \right.$

Interstitial compounds of transition metals are best dealt with in conjunction with alloys. They are conveniently classified (Hägg, 1931) according to the type of metal lattice and the coordination number of the non-metal atom. The latter atoms occupy holes in the rigid metal lattice,

and considerable variation in composition can occur. Many interstitial compounds are therefore examples of defect structures.

(c) Here we have to discuss a considerable number of inorganic structure types, and it is usual to subdivide these into ionic and homopolar structures. In the early days of the electronic theory of valency the distinction between ionic and homopolar bonds seemed very well defined, but the last few years have seen a definite change in this simple picture. We shall not consider here the various arguments for and against a clear-cut distinction between ionic and homopolar bonds, for it is more relevant to quote the following conclusions of Pauling regarding bond type (Pauling, 1939):

Bond.	Per cent. ionic character.
Alkali metals—non-metals	Essentially ionic, except Li—I, Li—C, Li—S (43 per cent. ionic).
Metals—O or F	Essentially ionic.
Be—O	63
B—O	44
Si—O	50

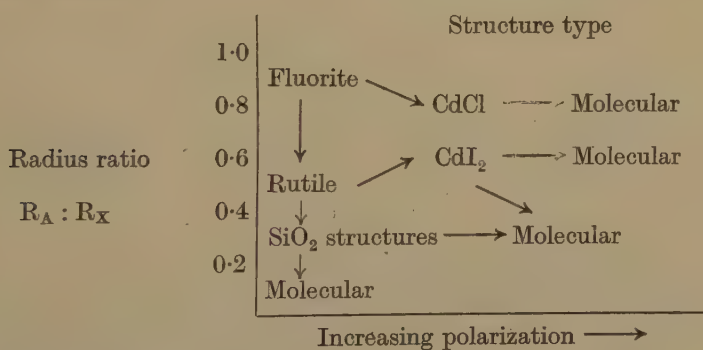
For the purpose of classifying structures we are not interested in the actual numerical values given in the table—we wish to emphasize the modern view that bonds can be of any intermediate kind between 100 per cent. ionic and 100 per cent. homopolar. A similar change in outlook is shown in the modern tendency to place less reliance on physical properties as criteria for deciding bond type. While it is true that certain physical properties may be associated with essentially ionic, and others with essentially homopolar bonds, in general it is dangerous to make deductions concerning bond type from physical properties. Perhaps the safest criteria are the atomic environments, *i. e.*, bond angles and inter-atomic distances, though two points must be borne in mind. Firstly, the same structure type may be possible for both extreme ionic and extreme homopolar bonds. This is the case when the disposition of the bonds from both atoms is the most symmetrical in three dimensions, *i. e.*, triangular, tetrahedral or octahedral, as opposed to three pyramidal, four square or six trigonal prism bonds. The zinc blende structure illustrates this point. Secondly, in different crystals with the same structure, associated in the classical treatment with a certain degree of polarization of one of the atoms (see below), there may be considerable differences in bond type. For example, the bonds Mg—Cl are probably far more ionic than Cd—Cl, but both MgCl_2 and CdCl_2 crystallize with the same structure. We must now consider the conventional treatment of simple inorganic crystals.

Radius Ratio and Polarization.

It is usual to describe ionic compounds according to their empirical formulæ and to attempt to account for the transitions from one structure type to another. In this connexion the concepts of radius ratio and polarization have been widely used, and the classical studies of Goldschmidt were of the greatest importance in the development of crystal chemistry as a science. The use of the radius ratio concept in determining the immediate environment of an ion implies firstly that the ion is a spherically symmetrically charged body, so that its neighbours in the crystal will be oriented in the most symmetrical way in three dimensions, viz.:—

	for 2 neighbours :	straight line
„ 3	„	: triangle
„ 4	„	: tetrahedron
„ 6	„	: octahedron
„ 8	„	: cube ;

and this appears to be the case in the essentially ionic structures. Secondly, it implies that the ion possesses a characteristic radius. The alteration in apparent ionic radius according to the number and charges of the neighbours necessitated the introduction of the conception of polarization, a general term to cover ionic distortion. In the limit the distortion of one ion by another leads to the formation of electronic orbits common to the two atoms, *i. e.*, the formation of a homopolar (covalent) bond. For compounds of empirical formula AX_2 the classical presentation due to Goldschmidt may be summarized in a scheme showing the change with increasing polarization from ionic to molecular structures.



Now the value of the radius ratio concept in accounting, for example, for the appearance of the fluorite and rutile structures is well known. However, a transition involving no change of coordination number

(*e.g.*, rutile to cadmium iodide) is attributed to the increasing polarization of one of the ions. In such a case the change from the triangular co-ordination of oxygen by titanium in rutile to the pyramidal binding of iodine to three cadmium atoms in CdI_2 indicates that the Cd—I bonds are less ionic than are the Ti—O bonds. It is therefore important to realize that the A—X bonds in these layer structures may have considerable covalent character, and to avoid statements of the kind: "Nevertheless the binding within the layers is purely ionic in character" (Evans, 'Crystal Chemistry,' p. 191). It would appear preferable to confine the use of the radius ratio criterion to essentially ionic structures and to abandon the use of diagrams of the kind shown above, in which the formation of layer and molecular structures is related to radius ratio and polarization. The implied extension of the radius ratio concept to semi-ionic and particularly to homopolar crystals is unsound. Its application to complex ions necessitates the adoption of ionic radii for highly charged cations such as C^{4+} , N^{5+} , S^{6+} , Mn^{7+} , etc., and is responsible for the extraordinary way in which these complex ions are sometimes described. After assuming that a group such as the SO_4 ion consists of a central S^{6+} ion surrounded by four O^{2-} ions, it is pointed out that the S—O bonds are partially (or even essentially) homopolar owing to the great polarizing power of the small, highly-charged S^{6+} ion. A further example of the unjustifiable extension of the radius ratio concept is found in the treatment of compounds with empirical formulæ ABX_3 , ABX_4 , and A_2BX_4 by Stillwell ('Crystal Chemistry.')

Crystals of these compounds are considered as aggregates of ions A, B, and X (even sulphates), and the following is taken from the table (Stillwell, p. 279) showing the relation of the radius ratio $r_A : r_X$ to the co-ordination number of A by X:—

$r_A : r_X$.	Coordination number of A.		
	Predicted.	Observed.	Structure.
0.29–0.48	4–6	4	Be_2SiO_4
0.28–0.48	4–6	8	K_2PtCl_4

The observed coordination of potassium by eight chlorine in K_2PtCl_4 cannot be explained in terms of the ratio of the radii of K^+ and Cl^- , for the good reason that there are no discrete Cl^- ions in the structure, the PtCl_4^{2-} ion being a covalent grouping, and Stillwell is obliged to conclude:—"The structure of the K_2PtCl_4 group is peculiar Pauling has shown that the unique directed properties of the B—X bonds brand

them as covalent rather than ionic bonds It is probably the abnormal planar distribution of the Cl ions which permits the coordination of eight by each A ion, rather than the six predicted by the radius ratio" (Stillwell, p. 280).

The discussion of the same compounds A_2BX_4 by Evans is considered in Appendix II.

Alternative Approach to Simple Binary Structures.

It is suggested that some of the difficulties mentioned above may be avoided in the following way. The relative dispositions in space of a number of covalent bonds may be deduced theoretically for a particular element. We may therefore draw up a table relating structures to the number and configuration of the bonds formed by the atoms A and X:—

TABLE I.

A bonds.	X bonds.			
	1.	2 non-linear.	3 pyramidal.	4 tetrahedral.
4 planar	Finite molecules AX_n	$PdCl_2$ (chains)		PtS
4 tetrahedral		SiS_2 (chains) HgI_2 (layers) GeS_2		ZnS ZnO
6 octahedral		$CrCl_3$ BiI_3 CrO_3 } (layers)	$CdCl_2$ CdI_2 } (layers)	(Al_2O_3)
6 trigonal prism			MoS_2 (layers)	

A structure in which the A—X bonds are essentially homopolar will be found in Table I., but it must be emphasized that the converse is not necessarily true, *i. e.*, the presence of a structure type in this table does not imply that it contains homopolar bonds. This point is amplified later. A similar table will contain structures in which the configurations of the bonds are the most symmetrical in three dimensions, *i. e.*, structures in which the bonds are essentially ionic. These tables are not complete, being intended only to illustrate general principles.

In this way we present a general picture of binary compounds in which many less common, but no less interesting, structures find a place. As an example we may cite the structure of SnS, which may be regarded as a

highly deformed NaCl structure in which three of the neighbours of an atom are much closer than the remaining three. One feature of this treatment may be noted here. Certain structure types, *e. g.*, zincblende and wurzite, appear in both tables. No information regarding bond type can therefore be obtained from the fact that a compound crystallizes with one of these structures, a point which is not made clear by some authors †.

TABLE II.

A bonds.	X bonds.			
	2 linear.	3 triangular.	4 tetrahedral.	6 octahedral.
4 tetrahedral			ZnS, ZnO	
6 octahedral	ReO ₃	TiO ₂	(Al ₂ O ₃)	NaCl (SnS)
8 cubic			CaF ₂	

This paper concludes with a critical account of the general classification of crystals according to bond type and of the subdivision of ionic crystals on the basis of electrostatic bond strength.

APPENDIX I.

The Classification of Crystal Structures according to Bond Type.

The early recognition of the four types of bonds in crystals suggested the allocation of crystals to one of four classes, *viz.*, ionic, homopolar, metallic and molecular. There are, however, comparatively few structures in which the bonds are all of the same type and are, in addition, pure bonds (*i. e.*, 100 per cent. ionic, etc.). We have, therefore, to recognize the following intermediate classes :—

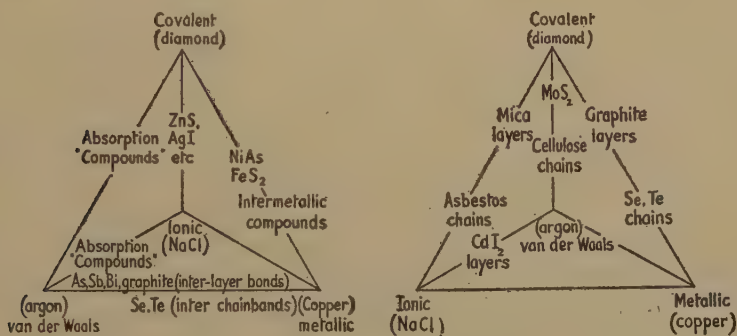
- (i.) All the bonds are similar, but are not pure types (*e. g.*, Si—O, 50 per cent. ionic, 50 per cent. homopolar).
- (ii.) The bonds are of two different kinds.
- (iii.) Combinations of (i.) and (ii.). For example, in CdCl₂ we may distinguish the Cd—Cl and the interlayer Cl—Cl bonds as being of

† Stillwell, for example, states : “ It may be said that the beryllium salts and MgTe are largely ionic *in spite of* their structure.” (Present author’s italics.)

different types, but in addition the Cd—Cl bonds are themselves intermediate between ionic and homopolar bonds.

It is instructive to see how these complications are dealt with in two recent books on crystal chemistry †. Both authors discuss the nature of the various bonds, and the evidence for and against a clear-cut distinction between ionic and homopolar bonds is set forth at considerable length by Stillwell. As a result, this author decides against an initial division into ionic and homopolar crystals and he does not suggest any comprehensive classification of structures. After dealing with binary compounds, however, he adopts Grimm's classification ‡ of these compounds "according to type of binding force." He gives two charts showing (1) intermediate bond types and (2) chain and layer structures. Of the "intermediate bond types" he says:—"The bonds are equivalent in all directions but a bond is intermediate in the sense that it is the average of an ionic and a covalent component."

The second chart, he says:—"Shows the relationship of structures with two or more co-existent bond types (as in SnI_4); the intermediate types are not representative of a gradual transition of bond type within the crystal, but are those in which one type of bond, predominating in two crystallographic directions, coexists with a second bond type predominating in the third direction."



This is clearly a compromise between a bond type classification and a geometrical one, and it will be seen that it does not afford a complete classification of crystal structures, even of binary compounds, for Stillwell adds:—"There appears to be no logical place on either of these charts for the simple molecular lattice, like CO_2 or SnI_4 . . ."

† 'Crystal Chemistry,' R. C. Evans, Univ. Press, Cambridge, 1939; 'Crystal Chemistry,' C. W. Stillwell, McGraw-Hill Book Co., 1938.

‡ Grimm, 1928.

TABLE I.

	Metallic.		Homopolar.		Ionic.			van der Waals.	
	Homo-desmic.	Hetero-desmic.	Homo-desmic.	Hetero-desmic.	Iso-desmic.	Meso-desmic.	Aniso-desmic.	Homo-desmic.	Hetero-desmic.
More or less close-packed †.									
Chain crystals †.									
Layer crystals †.									
Framework crystals †.	(1)					(2)			3)

† The terminology of these geometrical subdivisions is unfortunate. The description of diamond, ZnS, etc., as "more or less close-packed" crystals is difficult to understand. The author means "crystals containing 3-dimensional complexes."

(1), (2), (3). The placing of (1) interstitial compounds, (2) feldspars and zeolites, (3) globular proteins, in the same horizontal class is equally misleading. The framework silicates contain 3-dimensional and the globular proteins finite complexes.

There is, in fact, no place for any structure containing finite complexes—complex ions or molecules. Evans attempts to overcome this difficulty by distinguishing structures in which all the forces are of the same nature (homodesmic) from those in which forces of different types operate (heterodesmic). His broad classification then has secondary divisions according to the geometrical nature of the structure, as shown below. The subdivision of three of the vertical classes into the homo- and hetero-desmic divisions appears at first sight to be quite contradictory, but Evans explains (p. 41) that “heterodesmic compounds will be considered in each case under the heading of the weakest binding present.” However, this distinction between the two types of structures is not always retained, and for the detailed description of molecular crystals a new classification

TABLE II.

Molecular shape.	Type of intermolecular binding force.		
	Apolar.	Polar.	Ionic.
Simple symmetrical molecules.	Inert gases, O_2 , CH_4	Pentaerythritol	Oxalic acid, oxalates
Long molecules			
Flat molecules			
Complex 3-dimensional molecules			

is, in fact, adopted. Evans points out (p. 168) that organic salts, for example, should be included among ionic compounds, but he prefers to describe them under “molecular” compounds with the explanation:—“A convenient dividing line between ionic and molecular structures is the point at which the shape and polarity of the ion becomes of more importance than its size or charge.”

The scheme adopted for molecular structures is shown in Table II., which Evans describes as “an amplified presentation” of the van der Waals section of Table I. The examples shown are a selection of those given by Evans.

According to the author, oxalic acid and the oxalates are described under molecular structures instead of ionic because “the structures of these bodies are no longer determined by the packing of simple spherical ions or ionic groups, but by the packing of complex ions of irregular

shape" (p. 168), yet they are classed as "simple symmetrical molecules" in Table II. The secondary classification according to the weakest binding force (see above) is not consistent with the subdivision according to molecular shape. The structures of pentaerythritol and α -oxalic acid, for example, are determined by hydrogen bonds and are in some ways more closely related to layer structures than to structures containing simple symmetrical molecules. As soon as a classification takes account of molecular shape it must also recognize that intermolecular hydrogen bonds, being limited in number and acting between specific pairs of atoms of different molecules, can play as important a part in determining the structure and physical properties of the crystal as does the shape of the molecule itself. No allowance is made in Evans's classification for structures containing hydrogen bonds. Some of these structures are described in a separate chapter "Ionic Compounds containing Hydrogen," others under molecular crystals. Defect structures also are not included as part of the general scheme.

APPENDIX II.

A Classification of Ionic Crystals by Bond Strengths.

The concept of bond strength as defined by Pauling in formulating his "rules" for ionic structures has been of considerable utility. The later use by Bernal in defining a complex ion has now been extended by Evans to form the basis of a comprehensive classification of ionic crystals. It is important to examine critically any new classification of this sort, particularly since a new terminology may be easier to introduce than to eradicate. Evans defines the strength of a bond $A-X$ in a coordination group AX_n as the charge on the cation divided by n . Then if for any anion X^{z-} in the structure we consider the strength of the bond from X to an adjacent cation A , we may distinguish three cases according as the strength of the bond $A-X$ is greater than, equal to, or less than, $\frac{1}{2}z$. In this way ionic structures are subdivided into anisodesmic, mesodesmic, and isodesmic structures respectively. In anisodesmic structures "the single bond between X and A is stronger than all the remaining bonds linking X to the rest of the structure. In such a case the cation A and its several X neighbours form a discrete group in the structure." In mesodesmic structures, for a coordination group $A^aX_n^x$, $a/n = x/2$. The following abbreviated table illustrates the classification. It is important to note the implication of this nomenclature. The term anisodesmic is specifically designed to indicate that "the bonds differ only in degree," another term—"heterodesmic"—being applied to structures in which the bonds are "essentially different in kind" (Evans, p. 166, footnote). In other words it is assumed, for the purpose of this classification, that all

the bonds in, for example, CaCO_3 or K_2PdCl_4 , are ionic but differ only in electrostatic bond strength. Let us now examine this classification in more detail.

Isodesmic.	Mesodesmic.	Anisodesmic.		
		Univalent.	Divalent.	Trivalent.
F	Borates	CN	CO_3 SO_3	PO_3
Cl	Silicates	ClO_2		
Br	Germanates		SO_4 PdCl_4	AsO_3
I		ClO_3		
OH	AlF_6^{3-}		S_2O_3 S_2O_5	PO_4
O^{2-}	$\text{Cr}(\text{CN})_6^{3-}$	NO_3		

(a) *Anisodesmic Structures.*

The formulation of, for example, the carbonate and sulphate ions as $\text{C}^{4+}(\text{O}^{2-})_3^-$ and $\text{S}^{6+}(\text{O}^{2-})_4^-$ respectively naturally leads to difficulties concerning interatomic distances and the configurations of the groups. For example :—

Ions.	Sum of Crystal radii †.	Distance observed.
CO_3	1.55 A.	1.31 A.
SO_4	1.69	1.51
BO_3	1.60	1.35
SiO_4	1.81	1.60

(The values for the borate and orthosilicate ions, classed as mesodesmic, are referred to later.)

With reference to the short observed distances we find the following statement :—“ In all the complex ions occurring in anisodesmic structures the very strong polarizing power of the cation results in a pronounced tendency towards purely homopolar binding within the complex group, and this tendency is reflected in the structure by a considerable reduction in the interatomic distances compared with the values to be expected

† The ionic radii given here are taken from Pauling ('The Nature of the Chemical Bond'), since in spite of his classification Evans, in his table of ionic radii, gives no value for B^{3+} and only an upper limit for C^{4+} . The observed values of the distances are also the values adopted by Pauling (pp. 196, 222).

from a purely ionic linkage, and also in *occasional** departures in the shape of the group from the form to be expected on purely geometrical grounds" (Evans, p. 247).

The last part of this quotation should be emphasized since it refers to all non-linear AX_2 , pyramidal AX_3 and planar AX_4 groups, concerning the configuration of which Evans first quotes Zachariasen's rule, but goes on to say:—"Although Zachariasen's generalization is based on the assumption of a purely ionic binding in the groups discussed, it has been pointed out by Slater that the same conclusions can be reached when the *more valid picture*† of the partially homopolar nature of this binding is adopted."

In this classification compounds with similar empirical formulæ are necessarily discussed in different chapters and later considered as a group, and in reviewing the compounds A_2BX_4 , for example, Evans adopts the classical treatment of Goldschmidt relating the stability of the various structure types to the radii and polarizing powers of the A and B ions. No place in this scheme is suggested for the K_2PtCl_4 structure, which is classed as anisodesmic with the remarks:—"An ionic structure in which flat BX_4 groups occur is that of potassium chloroplatinite, K_2PtCl_4 . The flat form of the $PtCl_4$ groups must be interpreted as implying a purely homopolar binding within the groups" (Evans, p. 262).

With this may be compared the description of the K_2PdCl_4 structure by Stillwell already given (see p. 126).

We may sum up our conclusions regarding the "anisodesmic" group as follows:—If the shape of the group is that to be expected for ionic bonds, excessive polarization is invoked to account for interatomic distances. In other cases the bonds are frankly described as homopolar. This is sufficient to show the artificiality of classifying complex ions according to the strengths of the "ionic" bonds between the constituent atoms. It is relevant to note here Pauling's treatment of the planar ions CO_3 , BO_3 , and the tetrahedral ions SiO_4 , PO_4 , SO_4 , and ClO_4 . They are considered as essentially covalent complexes resonating between various forms involving single and double bonds. Pauling's comparison of the bond lengths S—O, etc., with the sum of the single bond covalent radii of the atoms, and the explanation of the smaller observed distances in terms of partial double bond character, should be contrasted with the ionic picture. Whether or not one agrees with the conclusions of a particular author it is important to realize that there are alternative ways of interpreting the observed facts.

† Present author's italics.

(b) Mesodesmic Structures.

Here again there is little to justify the distinction of certain groups on purely electrostatic grounds. The same difficulties arise with regard to interatomic distances as in the anisodesmic groups, as shown for the BO_3 and SiO_4 ions in the table above. A far more serious objection, however, is that groups such as $[\text{Co}(\text{CN})_6]^{3-}$ and $[\text{CoF}_6]^{3-}$ would both be classed as mesodesmic, whereas the observed magnetic moments of the potassium salts are as follows:— $\text{K}_3\text{Co}(\text{CN})_6$ 0.00, K_3CoF_6 5.3 Bohr magnetons, indicating that one complex is essentially covalent and the other essentially ionic. In the silicates, which are treated as ionic crystals, the mutual inclination of the oxygen bonds and the small Si—O distances indicate considerable homopolar character, and in both borates and silicates the binding is admitted to be partially homopolar (Evans, pp. 222 and 242). It is seen that the “mesodesmic” structures differ considerably among themselves in the nature of the interatomic binding, and that their separation as a class serves no useful purpose.

(c) Isodesmic Structures.

As already stated these are described in terms of radius, ratio, and polarization, the appearance of layer lattices being attributed to increasing polarization. The position of chain structures, such as SiS_2 and PdCl_2 , in this classification is doubtful. If classed as ionic compounds they would have to be described as mesodesmic, but actually the A—X bonds in these compounds are essentially homopolar. A different approach to the simple binary compounds—oxides, sulphides and halides—has already been suggested.

Conclusions.

We may summarize as follows. The idea of classifying inorganic structures according to the relative strengths of the bonds between different pairs of atoms is sound. It is not sound, however, to do this in terms of *electrostatic* bond strengths unless it is realized that the assignment of charges to the atoms is merely a device to facilitate classifying in terms of the relative values of a/n and $x/2$. It is clear from what has been said above that a rigid electrostatic treatment is neither generally acceptable nor capable of consistent development. The classification of Evans fails because it confuses bond strength with bond type. In order to recognize the carbonate group, for example, as a complex ion, it is not necessary to assume an ionic picture $\text{C}^{4+}\text{O}_3^{2-}$ and to show that $4/3 > 1$. The ion can be distinguished in crystals on the grounds of interatomic distances, which are the criteria adopted as the basis of the classification suggested in this paper.

Summary.

The presentation of systematic crystal chemistry demands a classification of crystal structures. It is suggested that the most satisfactory classification for this purpose is one which is primarily geometrical, and based as far as possible on observed interatomic distances. A classification of this kind is described. Appendices are devoted to critical reviews of (1) the classification of structures according to bond type, and (2) the subdivision of ionic structures according to electrostatic bond strength.

References.

- J. D. Bernal and D. Crowfoot, *Ann. Reps. Chem. Soc.* xxx. p. 407 (1934).
J. D. Bernal and H. D. Megaw, *Proc. Roy. Soc. Lond. A*, cli. p. 384 (1935).
C. W. Bunn, *Trans. Far. Soc.* xxxv. p. 482 (1939).
H. G. Grimm, *Naturwiss.* xxvii. p. 535 ; xxviii. p. 557 (1929).
G. Hägg, *Z. phys. Chem. B*, xii. p. 33 (1931).
C. Hermann, *Z. Krist.* lxxviii. p. 257 (1928) ; lxix. pp. 226, 250, 533 (1929).
F. Laves, *Z. Krist.* lxxiii. p. 202 (1930) ; *Z. f. Elektrochem.* xlv. p. 2 (1939).
P. Niggli, *Z. Krist.* lxxxvi. p. 65 (1933).
W. Nowacki, 'Homogene Raumteilung und Kristallstruktur,' Zürich (1935)
Z. Krist. c. p. 77 (1938).
L. Pauling, 'The Nature of the Chemical Bond,' Cornell Univ. Press (1939).
L. Pauling and L. O. Brockway, *Proc. Nat. Acad. Sci.* xx. p. 336 (1934).
L. Strock, *Z. Krist.* xciii. p. 285 (1936).
A. F. Wells, *Phil. Mag.* xxx. p. 103 (1940).

X. *Partial Derivatives of Derivatives.*

By W. S. KIMBALL, Ph.D., Robert College, Istanbul, Turkey *.

[Received May 23, 1939.]

PART II.—PARTIAL DERIVATIVES OF A LINE INTEGRAL IN THE CALCULUS OF VARIATIONS HAVING A CURVED PATH OF INTEGRATION.

Introduction.

THE method of differentials previously used to derive the new critical equations in the Calculus of Variations ⁽¹⁾ and the derivatives of line integrals by parallel displacement ⁽²⁾ is here extended to include second and higher partial derivatives of line integrals having a slope y' in the vector integrand, as in the Calculus of Variations proper. In contrast with Part I., where attention was restricted to such of these integrals as had a straight line for the path of integration at the upper and lower ends of the parallel displacement, we here treat integrals that have curved paths everywhere.

Formulas are accordingly derived for the partial derivatives of a derivative, or slope for a curved path. Two methods are employed, an analytical one. and a geometrical one of interpretation. The analytical one uses, in order to get the partials of y' with respect to y , elimination between the total derivative formulas $y''=y'_x+py'_y$, etc., and the partials of $y'=p$ with respect to x derived in Part I. for the straight-line path. These latter are taken over bodily and applied to curved paths, which they fit equally well because $dx=\Delta x$ in all cases, even though for curves Δy does not equal $dy=y'\Delta x$.

The geometrical method employs finite parallel displacements h and k just as previously ⁽²⁾ where no y' was present in the vector integrand. Successive differentiation with respect to h and k of the slope $y'(h, k)$ at the ends of the parallel displacement, and then setting $h=k=0$, gives all the partial derivatives of y' . Formulas are also given for the partials of the higher total derivatives y'' , y''' , etc., but no geometrical interpretation of these latter is here attempted.

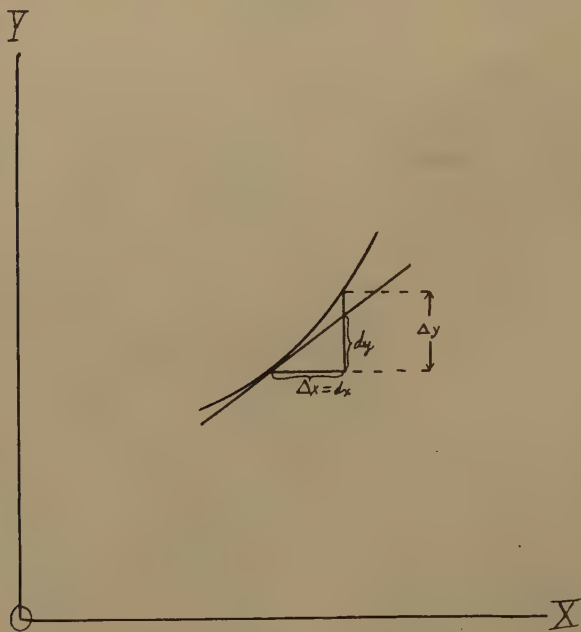
Rule I. previously derived ⁽²⁾ and Rule III. of Part I. for differentiation with respect to differentials are generalized into a single general rule, since they are shown to amount to practically the same thing. Besides

* Communicated by the Author.

the same sort of differentiation, they each provide changes for dependent variables corresponding to unit change in the independent variable of integration, *i. e.*, Rule I. by dropping the integral sign, and Rule III. by introducing $dx=n=1$ unit, according to the new definition of the derivative in Part I.

No attempt is here made to represent the derivatives of these integrals in terms of the curls of the vector integrand using Stokes's Theorem, but a way of doing so has been developed that checks the validity of the present treatment.

Fig. 1.



dy and Δy for a curved path.

§ 1. *The Partial Derivatives with Respect to X of the Slope y' of a Curved Path at (X, Y) are the same as if the Path were the Straight Tangent Line at (X, Y) .*

To obtain the partial derivatives of the slope y' of a curved path, we follow the established practice of treating $dx=\Delta x$ and $dy=y'\Delta x$, whereas Δy depends on the curve and the higher derivatives as indicated by fig. 1.

If the path of integration were the straight tangent line at the upper or lower end of a parallel displacement, then, in taking the partial derivatives of a line integral involving y' in the vector integrand, we would

have at these limits partials of y' which would be the same as those of p with respect to x and y given in Part I. We use throughout this article p as the slope of a straight line and y' as the slope of a curve. We now note the important fact, so far as partials with respect to x are concerned, that the partials of y' for a curve are always the same as those of p for the straight tangent line at the point (x, y) where the derivatives are taken. This follows at once from the established practice of treating $dx=\Delta x$ for a curve as well as a straight line, even though Δy does not equal $dy=y'\Delta x$, the value it would have in case the path were rectilinear. Thus, in partial differentiation with respect to x of a derivative by parallel displacements, we are in effect differentiating with respect to $dx=\Delta x$ partially and setting it afterwards equal to $n=1$ unit, and clearly it makes no difference whether the dy which remains constant is the $dy=\Delta y=p\Delta x$ of a straight line or the $dy=y'\Delta x$ (not $=\Delta y$) of the tangent line to a curve. Hence we take over bodily for a curve the formulas for the partials with respect to x as developed in Part I. for the slope p of a straight line :

$$\left. \begin{aligned} \frac{\partial y'}{\partial x} &= \frac{\partial y'}{\partial h}(0, 0) = y'_x = p_x = \pm \frac{y'}{n}, \\ \frac{\partial^2 y'}{\partial x^2} &= \frac{\partial^2 y'}{\partial h^2}(0, 0) = y'_{xx} = p_{xx} = \frac{2y'}{n^2}, \\ &\dots \dots \dots \\ \frac{\partial^m y'}{\partial x^m} &= (\pm 1)^m \frac{m!}{n^m} y' = \frac{\partial^m p}{\partial x^m}. \end{aligned} \right\} \begin{aligned} dx &= \Delta x = n = 1 \text{ unit,} \\ y' &= p = \frac{dy}{dx}. \end{aligned} \quad (1)$$

Eqs. (1) may be taken as a matter of definition as applied to a curve, but it is to be borne in mind that they are uniquely determined by the parallel displacement method of Part I. and previous articles, plus the established practice of putting on Δy all of the responsibility for departures from constancy in the ratio $\frac{\Delta y}{\Delta x}$ involved in departures of the path from rectilinearity.

To obtain the partials of y' with respect to y we have two methods available, an analytical method and a geometric one of interpretation. The analytic method is based on the familiar

$$\left. \begin{aligned} y'' &= y'_x + p y'_y = \frac{\partial y'}{\partial h}(0, 0) + p \frac{\partial y'}{\partial k}(0, 0), \\ y''' &= y''_x + p y''_y = \frac{\partial y''}{\partial h}(0, 0) + p \frac{\partial y''}{\partial k}(0, 0), \\ y^{iv} &= y'''_x + p y'''_y = \frac{\partial y'''}{\partial h}(0, 0) + p \frac{\partial y'''}{\partial k}(0, 0), \end{aligned} \right\} \begin{aligned} h &= k = 0, \\ p &= p_i \text{ or } p_u = \frac{dk}{dh}(0, 0). \end{aligned} \quad (2)$$

was given by adding $\mp h$ and $\mp k$ to the variables Δx and Δy according to the method of parallel displacements, and using $dx = \Delta x = n = 1$ unit, so as to give a unique result that would be the variation of a rate of change, according to the new definition for a derivative. So far as changes in x are concerned, we do exactly the same thing here, where all the blame for curvature is assumed by the increments of y .

This latter is exactly taken care of by replacing $dy_1(x_u, y_u) = p_u \Delta x = n p_u$ of the point (x_u, y_u) by $dy_1(X, y+k) = n y'(X, y+k)$, and using an increment q that differs from k in the same proportion. Thus, with upper and lower signs for the upper and lower ends of a parallel displacement, we have

$$y'(h, k) = \frac{dy_1(y+k) \mp q}{n \mp h} \equiv \frac{ny'(y+k) \left(1 \mp \frac{k}{np}\right)}{n \mp h}$$

$$\equiv y'(y+k) \pm \frac{y'(y+k)}{np} \left(\frac{hp-k}{1 \mp \frac{h}{n}}\right),$$

where

$$\frac{q}{k} = \frac{dy_1(y+k)}{dy_1(x, y)} = \frac{y'(y+k)}{p}, \text{ and } y'(y+k) \text{ means } y' \text{ at } (X, y+k).$$

. . . (5)

When k is zero, q is also zero, and $y' = p$ for the end point of the parallel displacement and (5) reduces to the varied slope (23) or (24) of the straight-line case with $k=0$ given in Part I. The slope $y'(y+k)$ appearing on the right of (5) is the slope of the original curve k units farther from the X -axis than (x_i, y_i) or (x_u, y_u) , at a point where coordinates for the upper

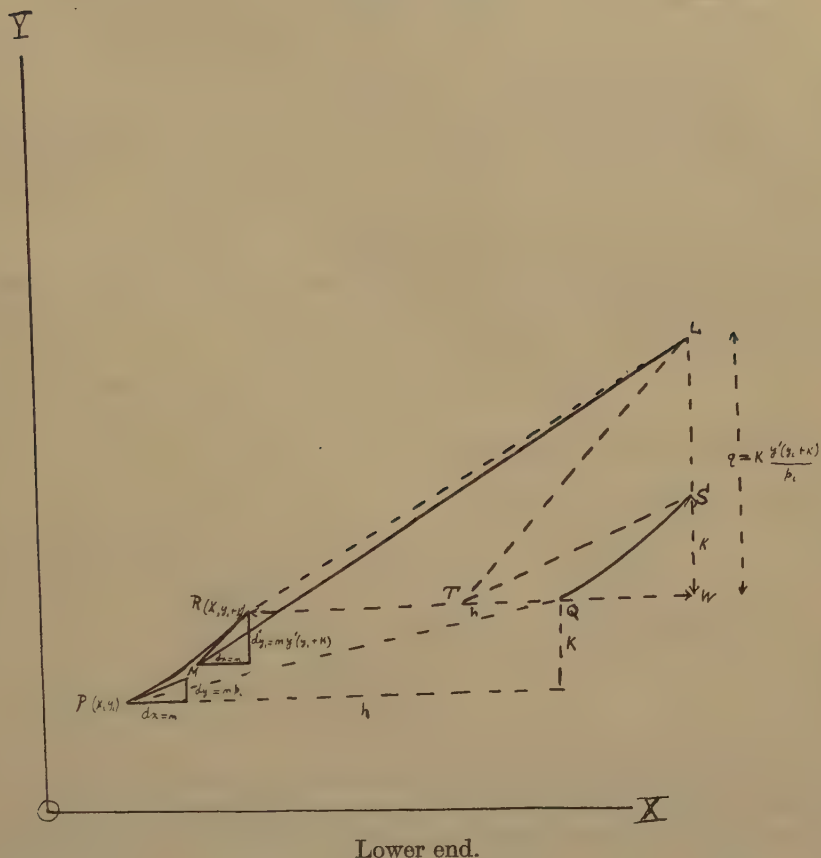
limit will be $X = \int_{y_u}^{y_u+k} dy/y'$ and y_u+k , such as shown by R in fig. 2.

This seems to be the simple and obvious way to vary the slope $y'(x_i, y_i)$ or $y'(x_u, y_u)$ at the ends of a parallel displacement due to increments h and k . Furthermore, it is unique in that the partial derivatives of (5) with respect to h and k at $h=k=0$ give all the partials of y' obtained by the analytic method of elimination used with (1) and (2) and its derivatives. This latter fact makes it unnecessary to concern ourselves with the possibility that the original curve might not ever reach a position k units away from the end point (x, y) . Thus (5) is to be viewed as an analytical device where h and k can be arbitrarily small, and assuming the existence of such derivatives as are needed for whatever results are yielded, it being understood that for simple curves the higher derivatives that come from the expansion of $y'(y+k)$ will mostly be zero.

Fig. 2 (a) shows the geometrical picture of the varied slope (5) at the upper end, and fig. 2 (b) shows it at the lower end of a parallel path displacement.

in the sense of being uniquely determined by choice of these two points (X, Y) and (x, y) once the three arguments, x, y , and p , of the end points are fixed upon; and (3) also shows that with (X, Y) fixed, the partial derivatives of (5) with respect to x and y are the same as those with respect to h and k respectively.

Fig. 2 (b).



The slope (5) is that of ML, obtained by combining altitudes and bases of the h, q triangle RWL with those of the unit base triangle at R.

§ 3. The Expansion of y' (h, k).

The expansion of (5) by Maclaurin's series in powers of k , and using the middle member of (5), is seen to be

$$\begin{aligned}
 y'(h, k) &= \frac{n}{n \mp h} \left[\left(y' + k \frac{dy'}{dy} + \frac{k^2}{2} \frac{d^2 y'}{dy^2} + \dots \right) \right. \\
 &\quad \left. \mp \frac{k}{np} \left(y' + k \frac{dy'}{dy} + \frac{k^2}{2} \frac{d^2 y'}{dy^2} + \dots \right) \right] \\
 &= \frac{n}{n \mp h} \left[\left(y' + \frac{k}{p} y'' + \frac{k^2}{2! p^2} y''' + \dots \right) \right. \\
 &\quad \left. \mp \frac{k}{np} \left(y' + \frac{k}{p} y'' + \frac{k^2}{2! p^2} y''' + \dots \right) \right] \\
 &= \frac{n}{n \mp h} \left[y' + k y'_y + \frac{k^2}{2!} y'_{yy} + \frac{k^3}{3!} y'_{yyy} + \dots \right].
 \end{aligned} \quad (6)$$

Here the second expansion comes from the first by the substitutions

$$\frac{y''}{p} = \frac{dy'}{dy}, \quad \frac{y'''}{p^2} = \frac{d^2 y'}{dy^2}, \quad \text{etc.,}$$

and the third merely collects the coefficients of each power of k and represents them by the indicated partials of y' with respect to y , whereby y'_y is seen to check (4) above and the higher partials are checked below. Now if we take

$$\frac{n}{n \mp h} = 1 \pm \frac{h}{n} + \frac{h^2}{n^2} \pm \frac{h^3}{n^3} + \dots$$

and substitute in the *middle* expansion (6), we have

$$\begin{aligned}
 y'(h, k) &= y' \pm \frac{y'}{np} (hp - k) + \frac{k}{p} y'' \\
 &\quad + \left[\frac{hy'}{n^2 p} (hp - k) \pm \frac{hk}{np} y'' + \frac{k^2}{2} \left(\frac{y'''}{p^2} \mp \frac{2y''}{np^2} \right) \right] + \dots \\
 &= y' + h y'_x + k y'_y + \frac{1}{2} (h^2 y'_{xx} + 2hk y'_{xy} + k^2 y'_{yy}) \\
 &\quad + \frac{1}{3!} (h^3 y'_{xxx} + 3h^2 k y'_{xxy} + \dots),
 \end{aligned} \quad (7)$$

where the partial derivatives are given by (1), (4), and those of y obtained below by the analytic method of elimination, using (1) and (2). Thus (5), with its associated geometrical interpretation, is seen by (7) to check the analytical derivations, which may also, of course, be obtained direct from (5) without any expansion, *i. e.*, take the partials of (5) with respect to h and k and then set $h=k=0$.

If we transpose y' in (7) to the left member and neglect second and higher powers of h and k , we have the differential or first variation of y' , using $\delta x = h$ and $\delta y = k$:

$$\delta y' = \frac{y'}{np} (hp - k) + k \frac{y''}{p} = \frac{y'}{np} \left(\frac{\delta x dy - \delta y dx}{dx} \right) + \delta y \frac{dy'}{dy} = \frac{y'}{np} \frac{d\sigma}{dx} + \delta y \frac{dy'}{dy}, \quad (8)$$

where

$$d\sigma = \delta x dy - \delta y dx,$$

which we note is the same as $(hp-k)/n$ for the straight-line case (29) in Part I., except for the second derivative term.

§ 4. *The Second Partial Derivatives of a Slope y' of a Curved Path at the Limits of a Parallel Displacement, and First Partial of y'' .*

As noted above, we may differentiate (5) with respect to h and k successively and then set $h=k=0$ to get the higher partials of y' . Also, so far as the x partials are concerned, we may differentiate here, as well as in Part I. (on account of (1)), with respect to the differential $dx=n$ and multiply by $(-1)^m$ at the upper limit, where m is the order of differentiation with respect to dx . Thus y'_{xy} may be obtained by differentiating (1), keeping n constant for variations in y , or again by differentiating (4), taking note that $\frac{1}{n} = \frac{1}{dx}$ is a common factor for the two terms of (4), and all else is constant for variations in x , so that it is as though we multiplied (4) by $\pm \frac{1}{n}$:

$$y'_{xy} = -\frac{y'}{n^2 p} \pm \frac{y''}{np} \dots \dots \dots (9)$$

We may now use (1), (2), (4), and (9) to obtain the partials of y'' . Thus differentiate the first of (2),

$$\left. \begin{aligned} y''_x &= y'_{xx} + p_x y'_y + p y'_{xy} = \frac{2y'}{n^2} - \frac{p}{n} \left(\frac{y'}{np} \mp \frac{y''}{p} \right) - p \left(\frac{y'}{n^2 p} \mp \frac{y''}{np} \right) = \pm \frac{2y''}{n}, \\ y''_x &= \mp \frac{\partial}{\partial x} \frac{dy}{dx^2} = \pm \frac{2d}{dx^3} dy = \pm \frac{2y''}{dx} = \pm \frac{2y''}{n}. \end{aligned} \right\} (10)$$

Note that the second line in (10) gives a second method of deriving y''_x , i. e. by differentiating with respect to the differential dx , and then setting it equal to 1 unit= n . The partial of y'' with respect to y is given by eliminating y''_x between (2) and (10):

$$y''_y = \frac{y'''}{p} \mp \frac{2y''}{np} \dots \dots \dots (11)$$

To obtain the second partial of y' with respect to y we have to use p_y of Part I., that differs from (4) in lacking the y'' term:

$$p_y = \mp \frac{1}{n} \dots \dots \dots (12)$$

Thus differentiating (4) and substituting (4), (11), and (12) gives

$$\left. \begin{aligned} y'_{yy} &= \frac{1}{p} \left(y''_{yy} \mp \frac{y'_y}{n} \right) - \frac{p_y}{p^2} \left(y'' \mp \frac{y'}{n} \right) = \frac{1}{p} \left(\frac{y'''}{p} \mp \frac{2y''}{np} \right) \\ &\quad \mp \frac{1}{np} \left(\frac{y''}{p} \mp \frac{y'}{np} \right) \pm \frac{1}{np^2} \left(y'' \mp \frac{y'}{n} \right), \quad \dots \quad (13) \\ y'_{yy} &= \frac{1}{p^2} \left(y''' \mp \frac{2y''}{n} \right) = \frac{y''_y}{p}, \quad \text{or} \quad y''_y = p y'_{yy}. \end{aligned} \right\}$$

Note that the power of n in the above denominators is the order of differentiation with respect to the differential $dx=n=1$ unit for the term in question. For odd powers of n the signs of the derivatives are reversed at upper and lower limits, but for even powers of n they are the same at both limits. Note also that the dimensions of each term of the second-order derivatives above (9), (10), (11) and (13) is L^{-2} , taking x , y and $n=dx=1$ unit of dimensions L , and hence y' of zero dimensions.

§ 5. *The Third Partial of the Slope y' , the Second Partial of y'' , and the First Partial Derivatives of y''' .*

We here derive those partial derivatives which are of dimensions L^{-3} . Thus

$$y''_{xx} = \mp \frac{\partial}{\partial x} \left(\pm 2 \frac{dy}{dx^3} \right) = 3! \frac{d^2 y}{dx^4} = 3! \frac{y''}{n^2} = \mp \frac{\partial}{\partial x} \left(\mp \frac{2y''}{n} \right) \quad \dots \quad (14)$$

is obtained by differentiating (10) with respect to dx and then setting $dx=n=1$ unit and multiplying by (-1) for the upper limit. The same result can be obtained by applying the quotient formula to the variables y'' and $dx=n$, and using (10), as indicated on the right. Likewise we may differentiate (10) with respect to y , using (11) and treating n as constant:

$$y''_{xy} = \pm \frac{2y'''}{np} - \frac{4y''}{n^2 p} \quad \dots \quad (15)$$

Now differentiate y''' given by (2), using (1), (11), (14), and (15):

$$y'''_x = y''_{xx} + y''_y p_x + p y''_{xy} = \pm \frac{3y'''}{n} = \mp \frac{\partial}{\partial x} \left(\frac{d^3 y}{dx^3} \right), \quad \dots \quad (16)$$

where again we note that (16), like (10) and (14), is obtainable by differentiating with respect to dx as on the right or by application of previous formulas. Now eliminate y'''_x between (2) and (16)

$$y'''_y = \frac{y^{iv}}{p} \mp \frac{3y'''}{np} \quad \dots \quad (17)$$

The third partials of y' are

$$\left. \begin{aligned} y'_{xxx} &= \pm 3! \frac{y'}{n^3}; & y'_{xxy} &= \frac{2}{n^2} y'_y = \frac{2}{n^2 p} \left(y'' \mp \frac{y'}{n} \right) = \frac{\partial^2}{\partial x^2} y'_y, \\ y'_{xyy} &= \mp \frac{\partial}{\partial x} y'_{yy} = \pm \frac{1}{n} y'_{yy} = \pm \frac{1}{n p^2} \left(y''' \mp \frac{2y''}{n} \right); \\ y'_{yyy} &= \frac{\partial}{\partial y} y'_{yy} = \frac{1}{p^3} \left(y^{iv} \mp \frac{3y'''}{n} \right) = \frac{1}{p^2} y''''; & \text{or } y'''_y &= p^2 y'_{yyy}. \end{aligned} \right\} \dots (18)$$

Note that here the partials of y'_y and y'_{yy} with respect to x are given by merely multiplying these by

$$\frac{2}{n^2} = n \frac{\partial^2}{\partial x^2} \left(\frac{1}{n} \right) \quad \text{and} \quad \pm \frac{1}{n} = \mp n \frac{\partial}{\partial x} \left(\frac{1}{n} \right).$$

This is because for these partials $\frac{1}{n} = \frac{1}{dx}$ comes into each as a common factor, with everything else constant:

$$y'_y = \frac{d^2 y}{n dy} \mp \frac{dy}{n dy} \quad \text{and} \quad y'_{yy} = \frac{d^3 y}{n dy^2} \mp 2 \frac{d^2 y}{n dy^2}.$$

In case of the last of (18), being the partial of (13) cancellation occurs, as when deriving (13). When we differentiate (11) or (13) with respect to y we have

$$y''_{yy} = \frac{1}{p^2} \left[y^{iv} \mp \frac{4y'''}{n} + \frac{2y''}{n^2} \right] \dots \dots \dots (19)$$

Eqs. (14) to (19) are the partials of dimensions L^{-3} , in which we note that odd powers of n (odd order partials with respect to dx) are associated with reversals of sign at upper and lower limit.

No attempt is here made to set up a geometrical picture like (5), giving y'' , y''' , etc., as functions of a finite pair (h, k) , but since we here provide a complete set of the derivatives of y'' , y''' , etc., with respect to k and h at $h=k=0$, the coefficients of the Maclaurin expansions are provided for, and the geometrical interpretation can very likely thus be obtained.

§ 6. *The true Derivatives are those at the Lower End of the Parallel Displacement. The General Formula for the Partial of y' , y'' , y''' , etc., in Terms of these Total Derivatives.*

Procedure similar to the above yields any number of higher partial derivatives

$$\left. \begin{aligned} \frac{\partial^4 y'}{\partial y^4} &= \frac{1}{p^4} \left(y^{iv} \mp \frac{4y^{iv}}{n} \right) = \frac{y^{iv}}{p^3}, \\ \frac{\partial^m y'}{\partial y^m} &= \frac{1}{p^m} \left(y^{(m+1)} \mp \frac{m y^{(m)}}{n} \right) = \frac{y^{(m)}}{p^{m-1}}, \end{aligned} \right\} y^{(m)} = \frac{d^m y}{dx^m} \dots \dots (20)$$

Also

$$\left. \begin{aligned} y_{xx}''' &= \frac{12y'''}{n^2}; \quad y_{xy}''' = \pm \frac{3}{np} \left(y^{iv} \mp \frac{3y'''}{n} \right), \\ y_{yy}''' &= \frac{1}{p^2} \left(y^v \mp \frac{6y^{iv}}{n} + \frac{6y'''}{n^2} \right), \\ y_x^{iv} &= \pm \frac{4y^{iv}}{n}; \quad y_y^{iv} = \frac{1}{p} \left(y^v \mp \frac{4y^{iv}}{n} \right). \end{aligned} \right\} D=L^{-4}. \quad (21)$$

$$\left. \begin{aligned} y_{xx}^{iv} &= \frac{20y^{iv}}{n^2}; \quad y_{xy}^{iv} = \pm \frac{4}{np} \left(y^v \mp \frac{4y^{iv}}{n} \right). \end{aligned} \right\} D=L^{-5}. \quad (22)$$

As always, these derivatives are of different sign at the upper and lower limits when the power of n is odd, but of the same sign when it is even, corresponding to odd and even order differentiation with respect to $dx=n=1$ unit.

All of the above derivatives may be obtained from a general formula,

$$\frac{\partial^m \psi^{(j)}}{\partial x^m} = (\pm 1)^m \frac{j(j+1) \dots (j+m-1)}{n^m} \psi^{(j)}, \quad (23)$$

where $\psi^{(j)}$ refers not only to the j th total derivative $\frac{d^j y}{dx^j} = y^{(j)}$ but to any one of its partials with respect to y . Thus if $j=1$, and hence $\psi^{(j)}=y'$, then (23) gives (1). If $\psi^{(j)}=p_y = \mp \frac{1}{n}$, with $j=1$, then (23) gives (34) of Part I., and for $j=2, 3, 4$, with $m=1$, (23) gives the first partials (10), (16), and (21), for $m=2$ it gives (14), (21), and (22) second partials.

The partials with respect to y are not obtained *directly* from (23), but by using (23) to eliminate y'_x, y''_x , etc., from (2), whereby y'_y, y''_y , etc., are expressible in terms of y', y'', y''' , etc.

But partials with respect to x of y'_y, y''_{yy}, y'''_y , etc., are all given by using these in (23) as illustrated by (9) and (18). The reason for this is that $y_y^{(j)}, y_{yy}^{(j)}$, etc., are like $y^{(j)}$ itself, all homogeneous functions of dx^{-j} , which under partial differentiation acts like a common factor, as pointed out in case of (18).

The basic reason for all this is that once we have differentiated an integral by Rule I., which removes integral signs, *all* the slope *variations* such as those of Part I. for the straight-line path, or (5) for the curved path, are outside the integral sign (under parallel displacement differentiation as herein), where the rôle of $dx=\Delta x$ is that of a finite variable for both straight line paths and curved paths and having a special important unit value $dx=n=1$ unit in derivative formulas so as to give *rates* of change, according to the new definition of the derivative. Formula (23), it will be seen, is nothing more than an application of Rule III. of Part I. for differentiation with respect to differentials, except that here its direct

application to second and higher partials with respect to y is denied instead of giving zero results as for the straight-line path of Part I. Thus all the derivatives of (23) are given by two steps that characterize Rule III. for differentiation with respect to differentials: (a) successive differentiation with respect to dx , treating it as a finite variable; (b) setting $dx=n=1$ unit, to give the derivative as a rate of change according to the new definition of the derivative. The third step whereby the factor (-1) is multiplied in at the *upper end only* of the parallel displacement each time we differentiate is not part of the differentiating process as such since these signs are outside h and k or h and q of (5), as pointed out in Part I. They came in at the upper end because the differentials function there as dependent variables under the restriction that is comprised in parallel displacement differentiation, whereby the upper end-piece PQ of fig. 2 *a* is the path of integration of a back-track integral from Q to P that helps to complete the closed-circuit path of integration.

The partials with respect to y are obtained by eliminating the above found x partials from the total derivatives, (2).

§ 7. *The General Rule for Differentiating with respect to Differentials that includes both working Rules I. and III.*

If we compare Rule I. for differentiation of a line integral with respect to differentials with Rule III., it will be found that they each amount to practically the same thing. In each case we differentiate with respect to differentials, treating them as ordinary finite variables because that is what they are; and furthermore, their increments are always in magnitude (for both straight lines and curves as per (3) above and Part I.) equal to the increments of the corresponding integral variables x and y at any place between the limits of a parallel displacement, and hence the resulting derivatives are the derivatives with respect to x and y when the proper sign is used, as well as suitable steps taken to make sure of using the required values of the dependent variable.

Formulas (37) and (38) of Part I. show that when we drop the integral sign, the integrands $f(x, y, y')$ and M and N at (x_i, y_i) become precisely the change in J per *unit* change in the variable of integration that we must use if we are to obtain derivatives or *rates* of change with respect to the variables, as shown by actually carrying out the integration as in (37) of Part I. Hence, according to Rule I., we are automatically supplied with the required value for the dependent variable by merely dropped the integral sign. Rule III obtains the same results by setting the differential of the independent variable of integration equal to one unit, $dx=n=1$ unit, whereby the variables dependent on it such as dJ , dS , and dy , take on the required values as shown in sections §§1 to 4 above in

Part I. Hence we have the generalization that includes both Rule I. and Rule III.

*General * Rule.*

The partial derivatives of a line integral with respect to x and y are given, so far as their differentials are concerned, by two steps: (a) differentiate with respect to the corresponding differential, using it as a finite variable; (b) make any needful changes that supply increments for dependent variables corresponding to *unit* increments for the independent variable of integration (*i. e.*, drop the integral sign as in Rule I. or set $dx=n=1$ unit as in Rule III.). The results are the true partial derivatives of J . These are those of J at the lower limit of the parallel displacement. At the upper limit (-1) has to be introduced for each derivative with respect to differentials according to Rule I. or Rule III., but not for derivatives with respect to the integral variables x and y .

Each application of Rule III. introduces $n=dx=1$ unit into the denominator of the resulting formula, so that the power of n in the denominator equals the order of differentiation according to Rule III. Differentiation by Rule I, however, does not introduce this dimensional factor $n=dx=1$ unit, since the equivalent result is attained by merely dropping the integral sign.

This General Rule depends for its validity upon the fact that $dx=\Delta x$ acts when outside the integral sign like a finite variable, with the same increments as x , and with a specially important and significant value $dx=\Delta x=n=1$ unit. Since $dx=\Delta x$ for both curves and straight line paths, this Rule for obtaining derivatives with respect to dx holds equally well in all cases. But for differentiation with respect to dy it holds completely for rectilinear paths where $dy=\Delta y$, and fails for second and higher partials (the first partials given by Rule I. has the straight tangent line for its path, as it were) with respect to y , using a *curved* integration path, where reliance must be put on the above derivative formulas obtained from (2) and (5).

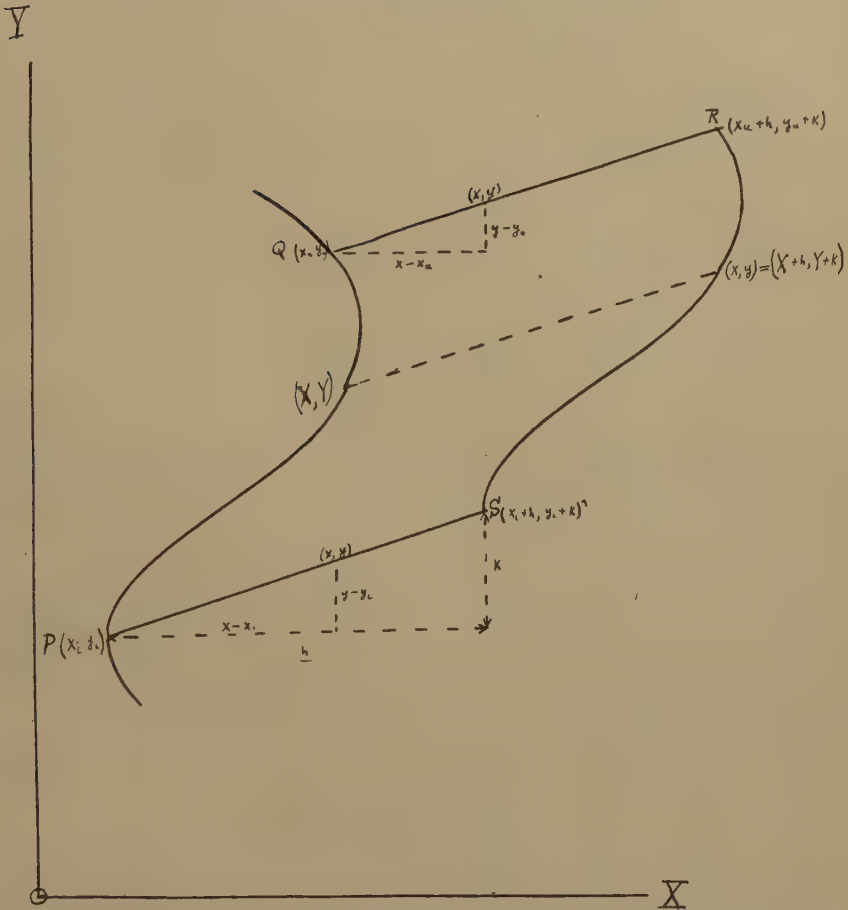
§ 8. *Second and Higher Partial Derivatives of Line Integrals in the Calculus of Variations.*

In case of a curved path of integration $y=\phi(x)$ satisfied by (X, Y) for a line integral in the Calculus of Variations, the varied integral due to parallel path displacements h and k will be taken along the three

* This rule is only employed *indirectly* for second and higher partials with respect to y obtained by eliminating the x partials from the total derivatives (2), a fact which is a consequence of the special role of x as an independent variable of integration with $dx=\Delta x$, whereas $dy \neq \Delta y$ except in case of a rectilinear path, for which the General Rule holds equally well for both x and y .

section paths PS, SR and RQ shown in fig. 3, where the upper and lower end pieces are straight lines of constant slope, $\frac{k}{h} = \frac{\delta y}{\delta x}$, and the displaced

Fig. 3.



The closed circuit PSRQ made by the parallel displaced path SR and straight line end pieces PS and Qk.

path section RS, has for its equation $y = \phi(x-h) + k$, being satisfied by $(x, y) = (X+h, Y+k)$. This varied integral may be written

$$J(h, k) = \int_{x_i y_i}^{x_u+h, y_u+k} (f - y' f_{y'}) \delta x + f_{y'} \delta y + \int_{x_i}^{x_u} f(x+h, y+k, y') dx - \int_{x_u y_u}^{x_u+h, y_u+k} (f - y' f_{y'}) \delta x + f_{y'} \delta y. \quad (24)$$

The third argument, y' , of the middle integral (24), which is taken along SR, is, of course, unchanged by *parallel* displacement, so that $M=f-y'f_{y'}$ and $N=f_{y'}$ can be combined in the usual way that results from treating y' as the slope of the path of integration. Along the end pieces, however, the third argument is given by (5), where h and k are replaced by $(x-x_i, y-y_i)$ and $(x-x_u, y-y_u)$ for the lower and upper end pieces, and since this third argument $y'(x-x_i, y-y_i)$, of (5) is not the same as $\frac{k}{h} = \frac{\delta y}{\delta x}$,

the slope of the path, we cannot combine M and N along the end pieces as done for the middle integral. The first partial derivatives of (24) with respect to h and k will be of the form

$$\left. \begin{aligned} \frac{\partial J_i^u}{\partial x} &= \frac{\partial J_i^u}{\partial h} = (f-y'f_{y'})_i + \int_{x_i}^{x_u} f_x(x+h, y+k, y') dx - (f-y'f_{y'})_u, \\ \frac{\partial J_i^u}{\partial y} &= \frac{\partial J_i^u}{\partial k} = f_{y'}(x_i+h, y_i+k, y'_i(h, k)) + \int_{x_i}^{x_u} f_{y'}(x+h, y+k, y') dx \\ &\quad - f_{y'}(x_u+h, y_u+k, y'_u(h, k)), \end{aligned} \right\} \quad (25)$$

where the above M and N now appear outside the integral signs, and the third argument is now in the form given by (5) as a function of h and k . By setting $h=k=0$ we obtain * the first partials by parallel displacement of the part of the line integral J between the indicated limits, and taken with respect to (x, y) any point on the path between the limits. Like the derivatives of an ordinary line integral, (25) is seen to be a *two-point function* of (X, Y) on the original path, and $(x, y)=(X+h, Y+k)$ its displaced position, as well as a function of h and k . Hence we may take as many derivatives of (25) with respect to h and k as we like, and by setting $h=k=0$ so obtain the corresponding derivatives with respect to x and y on the original curve.

The second partial derivatives so obtained are

$$\left. \begin{aligned} \frac{\partial^2 J_i^u}{\partial x^2} &= - \left[f_x - y' \left(f_{xy'} \pm \frac{y'}{n} f_{y'y'} \right) \right]_i^u + \int_{x_i}^{x_u} f_{xx} dx = \frac{\partial^2 J_i^u}{\partial h^2} (0, 0), \\ \frac{\partial}{\partial y} \frac{\partial J_i^u}{\partial x} &= - \left[f_y - y' \left(f_{yy'} + \frac{1}{p} \left(y'' \mp \frac{y'}{n} \right) f_{y'y'} \right) \right]_i^u + \int_{x_i}^{x_u} f_{xy} dx = \frac{\partial}{\partial k} \frac{\partial J_i^u}{\partial h} (0, 0), \\ \frac{\partial}{\partial x} \frac{\partial J_i^u}{\partial y} &= - \left[f_{xy'} \pm \frac{y'}{n} f_{y'y'} \right]_i^u + \int_{x_i}^{x_u} f_{xy} dx = \frac{\partial}{\partial h} \frac{\partial J_i^u}{\partial k} (0, 0), \\ \frac{\partial^2 J_i^u}{\partial y^2} &= - \left[f_{yy'} + \frac{1}{p} \left(y'' \mp \frac{y'}{n} \right) f_{y'y'} \right]_i^u + \int_{x_i}^{x_u} f_{yy} dx = \frac{\partial^2 J_i^u}{\partial k^2} (0, 0), \end{aligned} \right\} \quad \dots \quad (26)$$

* See eqs. (75), reference (2).

in which we note that the first and third above are the same as the corresponding derivatives in Part I. given by (45) using p_i and p_u for the straight-line paths since y'' does not enter into these two formulas. Likewise any number of higher derivatives are obtainable by successive differentiation. Eqs. (24), (25), and (26) illustrate the geometrical method of differentiating a line integral by parallel displacements, as in Part II. of the previous ⁽²⁾ article. The auxiliary variables h and k , need, however, never be introduced, being merely increments of x, y or their differentials. Thus (26) and higher derivatives are obtainable directly by differentiation with respect to x and y and their differentials, using Rule I. and Rule III. for the latter operations and taking note of the special formulas above for differentiation with respect to y .

Note that (26) will reduce to (15) of Part I., that refer to the *length* along a straight-line path, if we set $y'=p$, $y''=0$ and $f=\sqrt{1+p^2}$ and remember that the upper and lower signs refer to upper and lower limits respectively, whereby all terms of the first and last of (26) are positive, like those of (15) in Part I., and suggesting sufficiency conditions for a minimum of a function of two variables.

Since $y'=p$ always at every point on a path of integration, the second and third of (26) are always equal if Euler's equations are satisfied :

$$\frac{d}{dx}f_{y'}-f_y\equiv[f_{xy'}+y'_xf_{y'y'}]+[y'f_{yy'}+y'y'_yf_{y'y'}-f_y]=0, \quad (27)$$

because the brackets of (27) are the same as those of the second and third of (26) when the partials of y' are introduced, setting $y'=p$. On such an "extremal" arc given by (27) we thus have

$$M_y=f_y-y'\left(f_{yy'}+\frac{1}{p}\left(y''\mp\frac{y'}{n}\right)f_{y'y'}\right)=N_x=f_{xy'}\pm\frac{y'}{n}f_{y'y'}, \quad (28)$$

and hence, from (26),

$$\frac{\partial^2 J_i^u}{\partial y \partial x} = \frac{\partial^2 J_i^u}{\partial x \partial y} \quad (29)$$

A further point should not be lost sight of in connexion with (26). The expression outside the integral sign at the *lower* limit is the true partial derivative in question of J at the point (x_i, y_i) on the curve. Not true for the upper limit, because the (-1) introduced there with each application of Rule I. or Rule III. is extraneous to h and k , and the differentiation process as such, as pointed out in the General Rule of § 7, above, and in Part I.

§ 9. The Three Types of Derivative.

It is to be noted that in connexion with (26) there are *three* sets or types of derivative to be considered. Note clearly the distinctions between them, so as to avoid confusion :

(a) Partial derivatives (1) and (12) for the slope p of a *straight* tangent line at a point (x, y) on the curved path of integration $y=\phi(x)$.

(b) Partial derivatives (1), (4), etc., of the slope y' of the curved path of integration at one of the fixed limits (x_i, y_i, p_i) , (x_u, y_u, p_u) of a parallel displacement which may be measured by a finite h and k .

(c) Total derivatives along the path of integration $y=\phi(x)$, where p is identically equal to $y'=\frac{dy}{dx}$ and varies accordingly.

Types (a) and (b) are partial derivatives taken by parallel path displacements, whose differentials $(\delta x, \delta y)=(h, k)$ are the path "variations" of Euler and Lagrange used according to the method of parallel displacements, and refer to variations of *independent* variables. Here the coordinates of the end points do not change except for partial variations of the slope p referring to the straight tangent line at the end point. Type (c) has y dependent on x according to $y=\phi(x)$ and differentials related by $dy=\phi'(x)dx$. This last type may be employed to remove the integral signs from (26). Thus, set $(x_u, y_u, p_u)=(x, y, y')$, indicating that the upper limit may be taken as variable, and take the total derivative of both members, keeping $n=1$ unit=constant, as it, of course, would be, no matter what limits are taken for (26), and hence, also, for total variations of the limit coordinates (x_i, y_i, p_i) or (x_u, y_u, p_u) along the path.

As an illustration of the distinction between differentiation of Type (b) and of Type (c) above, we may take the derivative of, say, the first of (2) in the direction of the tangent to the curve. Thus for the total derivative of $y''=y'_x+py'_y$, and recalling that $\frac{dp}{dx}=0$, by (1) and (12), we have by the parallel displacement method

$$y'''=y'_{xx}+2py'_{xy}+p^2y'_{yy}. \quad \text{Type (2)} \quad (30)$$

and the total derivative along the path $y=\phi(x)$ is

$$\left. \begin{aligned} y''' &= \frac{d}{dx} y'_x + p \frac{d}{dx} y'_y = \frac{d}{dx} \left(\pm \frac{y'}{n} \right) + p \frac{d}{dx} \left(\frac{y''}{p} \mp \frac{y'}{np} \right), \\ &= \pm \frac{y''}{n} + p \left(\frac{y'''}{p} \mp \frac{y''}{np} \right) \equiv y'''. \end{aligned} \right\} \text{Type (3).} \quad (31)$$

Eq. (30) also will reduce to an identity when we introduce the partials given by (1), (9), and (13), being derivatives at a fixed limit point, by parallel path displacements. See § 4, above.

References.

- (1) W. S. Kimball, "New Basic Critical Equations and Methods in the Calculus of Variations," *Phil. Mag.* xxiii. p. 114 (1937).
- (2) W. S. Kimball, "Differentiation of Line Integrals by Parallel Displacement," *Phil. Mag.* xxv. p. 1 (1938).

XI. *On the Solution of certain Differential Equations of the Second Order.*

By L. ROTH, Imperial College of Science, London *

[Received March 5, 1940.]

I.

IN this paper we give a method of solving certain differential equations of the second order which occur in problems concerning oscillations in circuits †. The general equation of the type is represented by

$$Ay'' + By'^2 + Cy' + D = 0, \quad (1)$$

where A, B, C, D are real polynomials of assigned degrees in y , with constant coefficients. If, following Painlevé, we take (y, y', y'') for Cartesian co-ordinates in space, the equation (1) represents an algebraic surface of some order; it is thus suggested that such equations are to be classified according to the orders of their corresponding surfaces; and the usefulness of the classification is borne out in practice. We begin by considering equations of the quadric type.

1. *The Quadric Equation.*

This equation is of the form

$$(a_0y + a_1)y'' + b_0y'^2 + (c_0y + c_1)y' + d_0y^2 + d_1y + d_2 = 0. \quad (2)$$

To reduce this to a standard form, we first change the dependent variable from y to $a_0y + a_1$; then, grouping together all the terms which contain y as a factor, we may write the equation as

$$y(y'' + ay' + by + c) + f(y') = 0,$$

where $f(y')$ is a quadratic in y' only. If we now put

$$f(y') = l(y' - m)(y' - n),$$

where m and n are either real or conjugate complex quantities, the equation takes the form

$$y(y'' + ay' + by + c) + l(y' - m)(y' - n) = 0. \quad (3)$$

* Communicated by the Author.

† The outstanding problem in connexion with these equations is that of determining the periodic solutions, if any. We hope to consider this in a later paper. The present method of solution has, however, a certain interest of its own.

Interpreted geometrically, this form exhibits the two systems of generating lines on the representative quadratic surface. Proceeding in the same order of ideas, it is natural to replace (3) by the pair of simultaneous equations

$$y'' + ay' + by + c = \mu l(y' - m) \quad (4)$$

$$y' - n = -\mu y. \quad (5)$$

Here μ is a function of x —which we shall call the “integrating function”—such that (4) and (5) possess a common solution y .

If we eliminate y'' between (4) and (5), we can in general replace (4) and (5) by the equivalent equations

$$\{(l+1)\mu - a\}y' + (\mu' - b)y = lm\mu + c \quad (6)$$

$$y' + \mu y = n. \quad (7)$$

Solving for y' and y we obtain

$$y' : y : 1 = (\mu' - b)n - (lm\mu + c)\mu : [lm - (l+1)n]\mu + an + c : \mu' - (l+1)\mu^2 + a\mu - b. \quad (8)$$

It follows that in general the integrating function μ satisfies the second order equation

$$\frac{(\mu' - b)n - (lm\mu + c)\mu}{\mu' - (l+1)\mu^2 + a\mu - b} = \frac{d}{dx} \left\{ \frac{[lm - (l+1)n]\mu + an + c}{\mu' - (l+1)\mu^2 + a\mu - b} \right\}. \quad (9)$$

If any solution of (9) can be found, a corresponding value of y is then obtainable from (8) without further integration; and if the general solution of (9) is known, the general solution of (3) then follows, without integration, from (8).

Example.—Suppose that $b=0$ and that $(l+1)c + alm=0$. With these conditions it will be found that (9) is satisfied by $\mu = \frac{a}{2l+2} (1 + \tanh ax)$.

Then from (8) we have

$$y = \frac{[lm - (l+1)n] \frac{a}{2l+2} (1 + \tanh ax) + an + c}{\frac{a^2}{4l+4} \{2 \operatorname{sech}^2 ax - (1 + \tanh ax)^2 + 2(1 + \tanh ax)\}}$$

2. It is clear that, generally speaking, equation (9) will be considerably more complicated than the original equation (3). There is, however, one exceptional case in which the above procedure breaks down, and it is this case which yields to treatment. If the constants in (3) are subject to the relations

$$lm = (l+1)n, \quad an + c = 0, \quad (10)$$

we see from (8) that, unless y is to be everywhere zero, the equations (8) must fail to determine a value for y ; hence μ must satisfy the equation

$$\mu' - (l+1)\mu^2 + a\mu - b = 0. \quad (11)$$

With the conditions (10) and (11) it will be observed that, in (8), y' is likewise indeterminate: in fact, equations (6) and (7) have become identical. The general solution of (3) can now be obtained by solving (11), inserting the value of μ in (7), and performing a second integration. We thus have the general result:

The equation

$$y(y'' + ay' + by - an) + \frac{n}{m-n} (y' - m)(y' - n) = 0,$$

in which a, b, m, n are arbitrary parameters, can always be integrated completely*. As will now be shown, the general solution of the equation may take one of several distinct forms.

3. We assume in the first place that $l+1 \neq 0$; then, writing for convenience $l+1 = \rho$, we have from (11),

$$\mu' = \rho \left\{ \left(\mu - \frac{a}{2\rho} \right)^2 + \frac{4b\rho - a^2}{4\rho^2} \right\}.$$

We have now to distinguish three cases.

I. If $4b\rho - a^2 < 0$, we obtain, putting $\frac{a^2 - 4b\rho}{4} = D^2$, the solution

$$\mu = \frac{a}{2\rho} - \frac{D}{\rho} \tanh D(x+k),$$

where k is an arbitrary constant.

Returning to (7), we observe that the integrating factor of the equation is therefore $\exp \frac{a}{2\rho} (x+k) [\operatorname{sech} D(x+k)]^{1/\rho}$, so that y is given by the equation

$$\exp \frac{a}{2\rho} (x+k) [\operatorname{sech} D(x+k)]^{1/\rho} y = n \int \exp \frac{a}{2\rho} (x+k) [\operatorname{sech} D(x+k)]^{1/\rho} dx + \nu, \quad (12)$$

where ν is a constant of integration.

II. If $4b\rho - a^2 = 0$, we have, from (11),

$$\mu' = \rho \left(\mu - \frac{a}{2\rho} \right)^2.$$

* The success of the method depends upon the fact that we obtain a first order equation for the integrating function. It will be noticed that, for a real solution, m and n are necessarily real.

It follows that

$$\mu = \frac{a}{2\rho} - \frac{1}{\rho(x+k)}.$$

Hence, by (7), we obtain

$$\exp \frac{a}{2\rho} (x+k)[x+k]^{-1/\rho} y = n \int \exp \frac{a}{2\rho} (x+k)[x+k]^{-1/\rho} dx + \nu. \quad (13)$$

III. Finally, if $4b\rho - a^2 > 0$, we find that

$$\mu = \frac{a}{2\rho} + \frac{D}{\rho} \tan D(x+k),$$

where

$$\frac{4b\rho - a^2}{4} = D^2.$$

Hence, from (7), we obtain

$$\exp \frac{a}{2\rho} (x+k)[\sec D(x+k)]^{1/\rho} y = n \int \exp \frac{a}{2\rho} (x+k)[\sec D(x+k)]^{1/\rho} dx + \nu. \quad (14)$$

Among the above types are to be included equations in which the term in y'^2 is absent; such equations are of the form

$$y(y'' + ay' + by + c) - m(y' - n) = 0. \quad (15)$$

By a slight change in the previous notation we can at once apply the above results to equations of this type; and we find that (15) is integrable by the present method provided that

$$c + an = 0, \quad m = n.$$

The integrating function is given by the equation

$$\mu' = \mu^2 - a\mu + b,$$

so that the possible forms for y are obtainable by putting $\rho = 1$ in equations (12)–(14).

4. Consider now the case $l = -1$; evidently this implies that $m = 0$. Equation (11) becomes

$$\mu' + a\mu - b = 0.$$

Hence, if a is not zero, we have $\mu = e^{-a(x+k)} + b/a$, whence, by (7),

$$\exp \{e^{-a(x+k)} + b/a\} y = n \int \exp \{e^{-a(x+k)} + b/a\} dx + \nu. \quad (16)$$

If, however, a is zero, then $\mu = b(x+k)$, so that, by (7),

$$\exp \left\{ \frac{b}{2} (x+k)^2 \right\} y = n \int \exp \left\{ \frac{b}{2} (x+k)^2 \right\} dx + \nu. \quad (17)$$

5. *Examples.*—The various possibilities that may arise are illustrated in the following examples:—

Ex. 1. $4y(y''+y)-3(y'-m)(y'+3m)=0.$

Here $\mu=2 \tan \frac{1}{2}(x+k)$; then from (14) we obtain

$$y=-m\{2 \sin (x+k)+\sin (x+k) \cos (x+k)\}+\nu\{1+\cos (x+k)\}^2.$$

This result has a certain interest from another point of view; for if we take a finite Fourier series of the form

$$y=a_0+a_1 \cos x+a_2 \cos 2x+b_1 \sin x+b_2 \sin 2x,$$

as an assumed solution of a quadric equation of type (3), we find that the series must be of the form just obtained and, incidentally, that the equation itself must be of the above type.

Ex. 2. $y(y''-4y)-2(y'-m)(y'-2m)=0.$

Here $\mu=-2 \tan 2(x+k)$, whence, from (14), we have

$$y=m \tan 2(x+k)+\nu \sec 2(x+k).$$

Ex. 3. $y(y''+2y'+3y-4m)-2(y'-m)(y'-2m)=0.$

The equation for μ gives $\mu=-1+2 \tanh 2(x+k)$, whence, by (12), we have

$$e^{-(x+k)} \cosh 2(x+k) y=2 m \int e^{-(x+k)} \cosh 2(x+k) d x+\nu.$$

Thus
$$y=\frac{2 m}{3}\{1+2 \tanh 2(x+k)\}+\nu e^{(x+k)} \operatorname{sech} 2(x+k).$$

Ex. 4. $y(y''+2y'-y-4m)-2(y'-m)(y'-2m)=0.$

The equation for μ is $\mu'+(\mu+1)^2=0$, so that $\mu=-1+\frac{1}{x+k}$. Hence, by (13), we obtain

$$e^{-(x+k)}(x+k) y=2 m \int e^{-(x+k)}(x+k) d x+\nu,$$

whence

$$y=-2 m\left(1+\frac{1}{x+k}\right)+\frac{\nu e^{x+k}}{x+k}.$$

Ex. 5. $y(y''+2y'+2y-2m)-m(y'-m)=0.$ Here $\rho=1$, so that from (11) we have $\mu=1+\tan(x+k)$, whence, by (14),

$$e^x \sec (x+k) y=m \int e^x \sec (x+k) d x+\nu.$$

The integral may be evaluated by expanding $\sec (x+k)=[1-\sin ^2(x+k)]^{-\frac{1}{2}}$ as an infinite series and integrating term by term; the series is uniformly convergent except in the neighbourhood of points such that $|\sin (x+k)|=1$.

6. *Periodicity of the Solutions.*—It is clear from § 3 that when $4b\rho - a^2 > 0$, the given equation will in general possess periodic solutions. Thus, supposing, first, that a is not zero while $1/\rho$ is a negative integer, we can evaluate the integral in (14) in finite terms and then, taking $\nu=0$, divide through by the exponential factor, obtaining a solution with a real period. Again, if $1/\rho$ is not a negative integer, we may write $[\sec D(x+k)]^{1/\rho}$ in the form $[1 - \sin^2 D(x+k)]^{-1/2}$ and expand this function in an infinite series, uniformly convergent save in the neighbourhood of points where $|\sin D(x+k)| = 1$, and evaluate each of the integrals so obtained, again taking $\nu=0$.

If, however, $a=0$ and $n \neq 0$, we do not in general obtain periodic solutions: the integral $\int [\sec D(x+k)]^{1/\rho} dx$ represents a periodic function only if $1/\rho$ is a positive integer or an odd negative integer. When $a=0$ and $n=0$, all the solutions are periodic for any value of ρ .

The classification of the periodic solutions, according to the nature of their singularities, is shown in the following scheme:—

Analytic solutions: $a=0$, $n \neq 0$, $1/\rho$ an even positive integer.

$a=0$, $n=0$, $1/\rho$ a positive integer.

Meromorphic solutions: $a \neq 0$, $1/\rho$ a negative integer.

$a=0$, $n \neq 0$, $1/\rho$ an odd negative integer.

$a=0$, $n=0$, $1/\rho$ a negative integer.

Solutions with logarithmic singularities: $a=0$, $n \neq 0$, $1/\rho$ an odd positive integer.

Many-valued solutions with algebraic infinities: $a \neq 0$, $1/\rho$ rational, but not a negative integer.

$a=0$, $n=0$, $1/\rho$ rational non-integral.

(If ρ is positive, the solutions are finite.)

II.

7. *Equations whose First Integral is a Linear Equation.*—The above method of integrating certain types of quadric equation has been described in terms of the integrating function because this concept is useful for various purposes, other than the present. But there is another way of regarding the results which leads to an obvious extension of the method: we have in fact established conditions that the given equation (3) should have a first integral of the form

$$y' + \mu y = n.$$

Moreover, since μ is the solution of a first order equation, and therefore contains an arbitrary constant, it follows that the first integral so obtained

is the general one. As a first attempt to generalize this idea, let us see whether the quadric equation (2) can have a first integral of the form

$$y' = \mu y + \nu, \quad \dots \dots \dots (18)$$

in which μ and ν are functions of x . In that case we have, from (18),

$$y'' = \mu y' + \mu' y + \nu' = \mu(\mu y + \nu) + \mu' y + \nu' = (\mu' + \mu^2)y + \nu' + \mu\nu. \quad (19)$$

We now substitute for y' and y'' in (2) where, as usual, it is permissible to assume that $a_1 = 0$. We thus obtain a quadratic in y , namely

$$a_0 y \{ (\mu' + \mu^2)y + \nu' + \mu\nu \} + b_0 (\mu y + \nu)^2 + (c_0 y + c_1)(\mu y + \nu) + d_0 y^2 + d_1 y + d_2 = 0 \quad (20)$$

This becomes an identity if the coefficients of the various powers of y are each zero, that is, if

$$\begin{aligned} a_0(\mu' + \mu^2) + b_0\mu^2 + c_0\mu + d_0 &= 0 \\ a_0(\nu' + \mu\nu) + 2b_0\mu\nu + c_0\nu + c_1\mu + d_1 &= 0 \\ b_0\nu^2 + c_1\nu + d_2 &= 0 \end{aligned} \quad (21)$$

The third of these equations shows that ν must be a constant, and determines its value; for the equation can only vanish identically if $b_0 = c_1 = d_2 = 0$, in which case (2) reduces to a linear equation. The first of equations (21) determines μ , its solution containing an arbitrary constant; and the second equation, namely

$$(a_0 + 2b_0)\mu\nu + c_0\nu + c_1\mu + d_1 = 0,$$

has then to be satisfied for all values of μ . This means that

$$a_0\nu + 2b_0\nu + c_1 = 0, \text{ and } c_0\nu + d_1 = 0, \text{ where, by (21), } b_0\nu^2 + c_1\nu + d_2 = 0.$$

Thus two independent conditions are imposed on the constants; these are equivalent to the conditions established in § 2.

8. *Some Types of Cubic Equation.*—Consider now a cubic equation of the form (1), in which the polynomial B is a linear function of y ; if this is taken for dependent variable, the equation can be written as

$$(a_0 y^2 + a_1 y + a_2)y'' + b_0 y y'^2 + (c_0 y^2 + c_1 y + c_2)y' + d_0 y^3 + d_1 y^2 + d_2 y + d_3 = 0 \quad (22)$$

Suppose that this equation has a first integral of the form (18); then, substituting for y' and y'' from (18) and (19) in (22), and equating to zero the separate coefficients of the cubic polynomial in y so obtained, we derive the equations

$$a_0(\mu' + \mu^2) + b_0\mu^2 + c_0\mu + d_0 = 0. \quad (23.1)$$

$$a_0(\nu' + \mu\nu) + a_1(\mu' + \mu^2) + b_0\mu\nu + c_0\nu + c_1\mu + d_1 = 0. \quad (23.2)$$

$$a_1(\nu' + \mu\nu) + a_2(\mu' + \mu^2) + b_0\nu^2 + c_1\nu + c_2\mu + d_2 = 0. \quad (23.3)$$

$$a_2(\nu' + \mu\nu) + c_2\nu + d_3 = 0. \quad (23.4)$$

In general, equation (23.1) determines μ , the solution containing an arbitrary constant; and (23.4) then determines ν . The other two equations of condition have finally to be satisfied. Before, however, we proceed to a general discussion of these equations, it will be useful to consider a number of particular cases.

It is clear, first, that μ and ν can both be constants; in these circumstances we see that two independent conditions are imposed on the parameters in (22). But unless either μ or ν is arbitrary, the equation $y' = \mu y + \nu$ does not represent the general first integral of (22); on the other hand, it is seen by eliminating μ or ν from (18) that, if either is an arbitrary constant, the equation (22) is not of the cubic type. We shall, however, show that μ or ν can be a constant provided that the other is variable.

I. Suppose that ν alone is a constant; then from (23.4) we have

$$a_2 = 0, c_2\nu + d_3 = 0.$$

If μ is not constant, (23.3) must vanish identically, whence

$$a_1\nu + c_2 = 0, b_0\nu^2 + c_1\nu + d_2 = 0.$$

There are then two possibilities: either one of the equations (23.1), (23.2) vanishes identically or they reduce to the same equation. If (23.1) vanishes identically, $a_0 = b_0 = c_0 = d_0 = 0$. Thus (22) becomes an equation of quadric type; while if (23.2) vanishes identically,

$$a_1 = 0, (a_0 + 2b_0)\nu + c_1 = 0, c_0\nu + d_1 = 0.$$

From these conditions, together with those above, we deduce that $a_2 = c_2 = d_3 = 0$, so that once again (22) is a quadric equation.

If, instead, (23.1) and (23.2) are consistent, we require that

$$b_0 = 0, \frac{a_1}{a_0} = \frac{(a_0 + 2b_0)\nu + c_1}{c_0} = \frac{c_0\nu + d_1}{d_0}.$$

Thus six independent conditions are imposed on the parameters of the equation; these may be written in the form

$$\frac{c_2}{a_1} = \frac{d_2}{c_1} = \frac{d_3}{c_2} = \frac{a_0c_1 - a_1c_0}{a_0^2} = \frac{a_0d_1 - a_1d_0}{a_0c_0}, a_2 = 0, b_0 = 0.$$

II. Next, suppose that μ is a constant; its value is then determined by (23.1). Expressing the fact that the remaining three equations in ν are consistent, we have

$$\begin{aligned} a_0 : a_1 : a_2 &= (a_0 + 2b_0)\mu + c_0 : a_1\mu + c_1 : a_2\mu + c_2 \\ &= a_1\mu^2 + c_1\mu + d_1 : a_2\mu^2 + c_2\mu + d_2 : d_3. \end{aligned}$$

It follows that four independent conditions are then imposed on the parameters.

III. Consider now an equation for which $\nu = A\mu + B$, where A and B are constants. If we substitute for ν and ν' in (23.4) and identify the equation so obtained with (23.1), we have

$$b_0 = 0, \frac{a_2 A}{a_0} = \frac{c_2 A + a_2 B}{c_0} = \frac{c_2 B + d_3}{d_0}.$$

These relations determine A and B uniquely. Similarly, if we substitute for ν and ν' in (23.2) and (23.3) and identify these equations with (23.1), we obtain four relations, linear in A and B . Thus five independent conditions in all are imposed.

IV. After these preliminaries we are in a position to deal with the general case. Suppose now that neither μ nor ν is a constant: substituting for μ' and ν' from (23.1) and (23.4) in (23.2) and (23.3), we obtain the identities

$$\frac{a_1}{a_0} (b_0 \mu^2 + c_0 \mu + d_0) + \frac{a_0}{a_2} (c_2 \nu + d_3) = 2b_0 \mu \nu + c_0 \nu + c_1 \mu + d_1. \quad (24.1)$$

$$\frac{a_2}{a_0} (b_0 \mu^2 + c_0 \mu + d_0) + \frac{a_1}{a_2} (c_2 \nu + d_3) = b_0 \nu^2 + c_1 \nu + c_2 \mu + d_2. \quad (24.2)$$

If $b_0 = 0$, these are linear in μ and ν . There are then two possibilities: either ν is a linear function of μ , as in case III. above, or we have, from (24.1),

$$a_1 c_0 = a_0 c_1, \quad a_0 c_2 = a_2 c_0, \quad a_1 a_2 d_0 + a_0^2 d_3 = a_0 a_2 d_1,$$

with a similar set of conditions from (24.2). Thus six conditions are entailed in all. If, however, b_0 is not zero, we may solve (24.1) for ν and substitute the result in (24.2), obtaining a quartic in μ which must vanish identically. Now the coefficient of μ^4 in this quartic will be seen to be $b_0^3 (a_1^2 - 4a_0 a_2)$; hence we must have $a_1^2 = 4a_0 a_2$. Then, if we change

the dependent variable in (22) to $y + \sqrt{\frac{a_2}{a_0}}$, the first two terms of the

equation take the form $a_0 y^2 y'' + (b_0 y + b_1) y'^2$; it follows that, if we substitute as before from (18) and (19) and equate to zero the term independent of y , we find that ν must be a constant, ν_0 , say. Hence, in terms

of the original variable y , the first integral is $y' = \mu \left(y + \sqrt{\frac{a_2}{a_0}} \right) + \nu_0$; that

is, ν is a linear function of μ , as in case III., whence b_0 is necessarily zero.

The results of the above investigation may be summarized as follows:—

If a cubic equation of type (22) has a first integral of the form $y' = \mu y + \nu$, then either μ is a constant or else b_0 is zero; and in all cases the general solution of the equation is obtainable by quadratures.

9. Precisely similar methods may be applied to a cubic equation of the form

$$(a_0y^2+a_1y+a_2)y''+b_1y'^2+(c_0y^2+c_1y+c_2)y'+d_0y^3+d_1y^2+d_2y+d_3=0. \quad (25)$$

The conditions that this should have a first integral of type (18) are

$$a_0(\mu'+\mu^2)+c_0\mu+d_0=0. \quad (26.1)$$

$$a_0(\nu'+\mu\nu)+a_1(\mu'+\mu^2)+b_1\mu^2+c_0\nu+c_2\mu+d_1=0. \quad (26.2)$$

$$a_1(\nu'+\mu\nu)+a_2(\mu'+\mu^2)+2b_1\mu\nu+c_1\nu+c_2\mu+d_2=0. \quad (26.3)$$

$$a_2(\nu'+\mu\nu)+b_1\nu^2+c_2\nu+d_3=0. \quad (26.4)$$

Here again, if b_1 is not zero, we require μ to be a constant, its value being given by (26.1). The remaining equations can be consistent only if (26.2) and (26.3) vanish identically, that is, if

$$a_0=0, \quad c_0=0, \quad a_1=0,$$

and

$$b_1\mu^2+c_1\mu+d_1=0, \quad 2b_1\mu+c_1=0, \quad a_2\mu^2+c_2\mu+d_2=0.$$

Thus six independent conditions are entailed.

10. In conclusion, we may remark that such relations as the above, which must be satisfied by the parameters of the given equation in order to obtain a first integral of the desired type, are not likely to occur in practice; for this reason, any examples which we might give as illustrations of the method would necessarily be of a rather artificial character. On the other hand, the results have a natural application to equations for which these relations are satisfied *approximately*. We hope to consider such equations in a subsequent paper.

XII. *Velocity of Sound in some Binary Alloys.*

By GEORGE E. ALLAN, D.Sc.*

[Received May 12, 1941.]

IN this paper are given values of the velocity of sound in certain alloys. The observations deal with the binary alloys of tin-lead and of copper-zinc, and the results receive some confirmation from the measurements by other investigators on zinc-tin and on iron-nickel. Graphs are drawn to show the manner in which the velocity of sound changes as the composition of the alloy is varied.

Velocity of Sound in Tin-Lead Alloys.

Three rods of solder were procured which contained, nominally, 25, 50, and 75 per cent. by weight of tin. The rods when analysed from turnings taken from the ends were found to contain 22.3, 49.7, and 78.1 per cent. of tin, the rest being lead with a trace of iron. The rods, after being turned into cylindrical form, had lengths 33.69, 28.53, and 28.40 cm. and diameters 1.071, 1.054, and 1.131 cm. respectively. The tone heard when the rods were held at the middle between finger and thumb and struck with a hammer at the end, setting the rod vibrating transversely, was the second partial of the transverse series and the longitudinal tone sought was inaudible. The fundamental tone due to lateral vibrations was therefore employed in order to calculate the high frequency due to longitudinal vibrations, and the rods were examined in the following way:— Each rod was held at one of the two nodes about $\frac{2}{9}$ ths (0.2242) of the length of the rod from either end, and struck at the middle. The frequency of the note, N_T , thus obtained was measured by means of the monochord⁽¹⁾, and that for longitudinal vibrations, N_L , was calculated from Poisson's formula, which had been verified experimentally by Savart⁽²⁾,

$$N_L = \frac{L}{1.780 D} N_T$$

* Communicated by the Author.

where L is the length of the rod and D its diameter. The results thus obtained are given in the following table :—

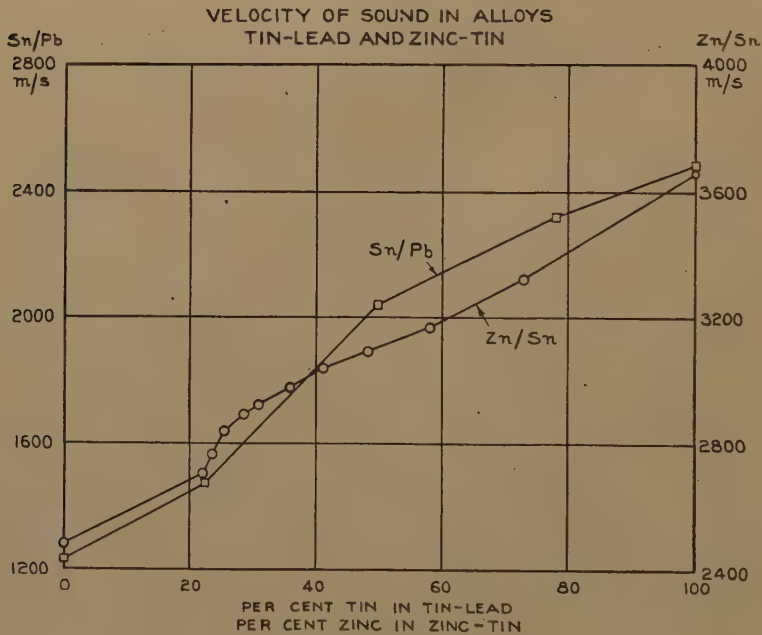
TABLE I.
Velocity of Sound in Tin-Lead Alloys.

Rod.	Per cent. tin.	N_T .	N_L .	Velocity of sound.
(Lead) ...	0 .	—	—	*1230 m/s.
Alloy 1...	22.3	124.4	2199	1481
2...	49.7	237.8	3616	2063
3...	78.1	291.4	4111	2334
(Tin).....	100	—	—	*2483

* from tables.

The course of the change may be followed in the graph for tin-lead.

Graph 1.



Gerosa's Measurements with Zinc-Tin Alloys.

A change similar in some respects to that described has been shown ⁽³⁾ by G. Gerosa to take place in the series of zinc-tin alloys. Gerosa investigated the velocity of sound in ten specimens in which one gramme-

molecule of zinc was alloyed with $1/5$, $2/5$, $3/5$ $10/5$ gramme-molecule of tin respectively. These were, with the pure metals, drawn into wires, and their frequencies were determined for longitudinal vibrations. The wires, 149.88 cm. long between the supports, were sounded, and a sonometer wire, vibrating transversely, was tuned to the note of the wire under test, of frequency say n . Let the length of the sonometer wire be l cm. For comparison a length 54.42 cm. of the sonometer wire gave a note in unison with a standard Koenig 512 fork, *i. e.*, one of frequency 256 cycles or double vibrations per second. Assuming that, for the two sonometer notes, frequency $\times$ length is constant,

$$n \times l = 256 \times 54.42 = 13932.$$

Thus, when l is known, $n = 13932/l$. For the pure zinc wire l was 11.38 cm. Its frequency was therefore 1224, and the velocity of sound in zinc is 3670 metre/sec. Similarly the frequency of the tin wire was 828 and the velocity of sound in tin 2483 m/s. (see Table I.).

If the proportions of zinc and tin are expressed in per cent. weight of zinc in each alloy, and the actual velocities given instead of their comparison with the velocity of sound in air, the following figures are obtained :—

TABLE II.

Gerosa's Results for Velocity of Sound in Zinc-Tin Alloys.

Metal.	Per cent. zinc.	Sonometer length.	Frequency.	Velocity of sound.
		l cm.		m/s.
Pure tin . . .	0	16.82	828	2483
Alloy 1.	21.6	15.47	901	2700
2.	23.4	15.07	925	2771
3.	25.6	14.73	946	2835
4.	28.2	14.47	963	2886
5.	31.5	14.27	976	2926
6.	35.5	14.06	991	2970
7.	40.8	13.83	1007	3020
8.	47.9	13.53	1030	3087
9.	57.9	13.14	1060	3178
10.	73.4	12.57	1108	3322
Pure zinc ..	100	11.38	1224	3670

The graph for zinc-tin alloys has been drawn from the above figures. It may be seen that there is, in the two curves, a similarity of shape or course, a rough proportionality between content of the first-mentioned metal of the pair and the change of velocity, and an almost equal increase

in the velocity of sound between the initial and final velocities for each pair of metals.

Velocity of Sound in Copper-Zinc Alloys.

Five long rods of copper alloyed with zinc were obtained which contained 53, 60, 70, 80, and 90 per cent. by weight of copper, and for these the length, diameter *, density, frequency and velocity of sound were measured. Each rod was clamped at its middle in a horizontal position within a narrow hard-leather ring on a retort stand. It was struck on the end with a hammer to elicit the fundamental tone of the rod due to longitudinal vibrations. The frequency of the tone was determined, as in the previous case, by means of the monochord, and the values found are given in the following table, which includes also the velocities of sound in the pure metals.

TABLE III.

Velocity of Sound in Copper-Zinc Alloys.

Rod.	Length cm.	Per. cent. copper.	Frequency.	Velocity of sound m/s.	Density gm./c.c.
(Zinc) ...	—	0	—	3670†	—
Alloy 1...	351.72	53	455.4	3204	8.355
2...	326.91	60	526.5	3442	8.421
3...	338.92	70	538.1	3647	8.569
4...	291.27	80	628.5	3661	8.639
5...	271.59	90	684.1	3716	8.784
Copper ..	182.97	99.9	1050	3846 ⁽¹⁾	8.928

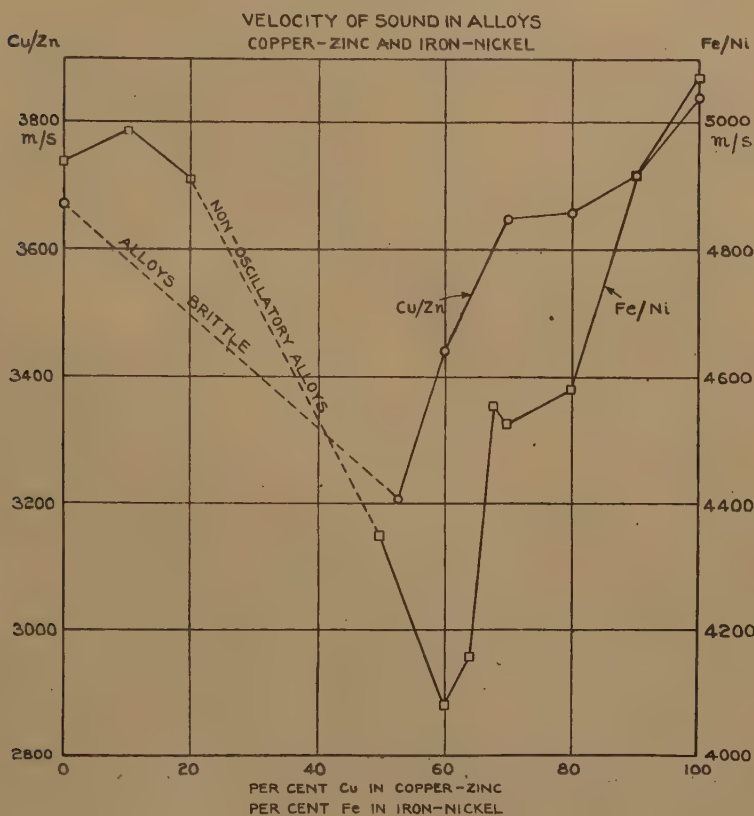
† Gerosa's result as recalculated.

The change in the velocity of sound in these brasses is shown in the second graph, which gives also the corresponding change in the velocity of sound in iron-nickel alloys as found by Pierce ⁽⁴⁾, using a magnetostriction method. A comparison of the two graphs brings out several points of resemblance. For instance, at low percentages of copper there is a region of great brittleness which makes the production and life of long rods a precarious process. In the same region the iron-nickel alloy is non-oscillatory. In both cases there is a big drop in the velocity of sound to a minimum value. Wertheim's values exhibit the same feature, and the chief effect of the impurities in his material appears to be

* The mean diameters of the brass rods were 1.301, 1.294, 1.300, 1.307, and 1.300 respectively.

an over-all lowering of the velocity of sound. As the velocity of sound increases, with copper content in one case and with iron content in the other, there is a pause at about 70 per cent. to 80 per cent., after which the increase of velocity proceeds at a greater rate. These points of resemblance seem to indicate a similar progressive change in the structure of the alloys.

Graph 2.



Intermetallic Compounds of Copper and Zinc.

Metallurgists are not agreed as to the number and composition of these, but such compounds are regarded as occurring in the region containing 0 to 53 per cent. of copper. The alloy which possesses maximum hardness (Cu_2Zn_3) contains 39.3 per cent. of copper⁽⁵⁾. The brasses of this region are referred to as γ -brasses, and their brittleness is ascribed to their content of γ -solution. Brasses with 53 to 64 per cent. of copper contain two kinds of grains and are known as $\alpha\beta$ -brasses. The $\alpha\beta$ -brasses

have a much greater tensile strength than those which contain more than 64 per cent. of copper, and which are known as α -brasses ⁽⁶⁾.

Velocity of Sound in the Elementary Metals.

The mathematical relation of the velocity of sound in a material to the elasticity and density of the material is well known. Hitherto, no attempt has been made, with one exception, to relate this physical constant to the molecular structure of the substance. M. R. Rao ⁽⁷⁾ has shown that, for a large number of liquids, the equation $v^3V = \text{constant}$ holds good, v being the velocity of sound in the liquid and V the molecular volume. The equation does not hold for solid elementary metals. It is the case, however, that for a considerable number of such metals the velocity of sound in them is greater, the smaller the atomic weight of the element. This may be verified by examining a table of atomic weights of elementary metals for which the velocity of sound is known. It may also be verified that there are numerous exceptions to this simple rule; and a law which could include the metallic elements would most probably have to take account of the position of the element on the periodic table. The simple property holds, however, for the metallic elements dealt with in this paper, and has been employed in the drawing of the graphs to give the change in the velocity of sound as the composition of the alloy is varied from 0 to 100 per cent. of the constituent with the smaller atomic weight.

The observations recorded in this paper were made in the Applied Physics Department of Glasgow University, and I have pleasure in expressing my thanks to Dr. John A. Mair for his analysis of the rods of solder, to Dr. Ernest W. Geyer for making enlarged copies of the graphs, and to Prof. Robert Hay, Glasgow Royal Technical College, for providing brass rods of definite chemical composition.

Literature.

- (1) G. E. Allan, *Phil. Mag.* (7) xxvi. p. 609 (1938).
- (2) Violle, 'Cours de Physique,' vol. ii. part 1, p. 207.
- (3) G. Gerosa, *Acad. d. Linc. Rend.* 4, iv. p. 127 (1888).
- (4) G. W. Pierce, *Am. Acad. Arts and Sciences*, lxiii. p. 1 (1927-28).
- (5) C. H. Desch, 'Intermetallic Compounds,' Longmans Green & Co. (1914).
- (6) U. R. Evans, 'Alloys,' vol. iv. pp. 57, 218.
- (7) M. R. Rao, *Indian Journ. Phys.* xiv. p. 109 (Apr. 1940).

XIII. *Biot-Savart Law and Newton's Third Law of Motion.*

By S. B. L. MATHUR, M.Sc. (Lecturer, University of Lucknow) *.

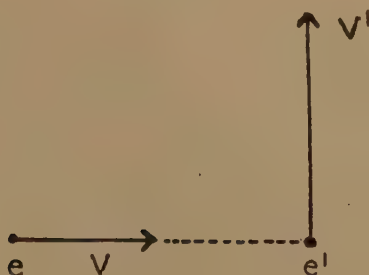
[Received October 17, 1939.]

It is not often realized that Newton's Law of Action and Reaction is not generally valid for the mutual forces between current elements (—the obvious case of parallel current elements being an exception—). This inapplicability of Newton's Law is not usually stated in the current text-books on electrodynamics, and a failure to recognize this fact is often a source of confusion in electrical calculations. It is the purpose of this note to discuss one or two examples which show how the application of Newton's third law in such cases leads to wrong results. The correct results are obtained when it is recognized that the third law is not applicable to mutual forces between current elements.

1. *Two Moving Point Charges.*

Let us first take the case of two point charges e and e' (fig. 1). If e has a velocity V directly towards e' , while e' has a velocity V' at right

Fig. 1.



angles to the line ee' , then, the Lorentz force on e' is zero, while the force on e is easily seen to be

$$\frac{ee'VV'}{c^2r^2},$$

where r is the distance between e , e' , and c the velocity of light. The two forces are thus neither equal in magnitude nor opposite in direction.

* Communicated by Dr. D. S. Kothari.

In fact, if in general ⁽¹⁾ F be the Lorentz force on e , and F' on e' , then, in vector notation,

$$F + F' = -\frac{ee'}{c^2} \nabla_e \frac{1}{r} (V \times V').$$

Newton's third law is, therefore, only true when this expression vanishes, i. e., either V and V' are parallel or both V and V' are at right angles to r .

2. Dunton's Circuit-breaker.

We now consider another example.

Fig. 2 illustrates an experimental circuit-breaker arrangement discussed by Dunton ⁽²⁾. A current flows in *via* ABCD. M and N are mercury cups where the arcs would be formed when the circuit-breaker begins to open. MADN is fixed, while the portion MBCN is free to move. The dimensions of the circuit are shown in the figure: $BC = 60r$, $AB = 180r$, $BM = 60r$, where r is the radius of the conductor of which the circuit is made.

Let us first assume that Newton's Law of Action and Reaction applies to the mutual forces between current elements, and calculate on this assumption the outward force acting on MBCN. This assumption means that in calculating the force on MBCN the mutual forces between the conductors composing it (BC, MB, CN) are to be ignored, and the effect on MBCN will be due solely to the remaining (fixed) circuit MADN. The outward force exerted by MADN and MBCN can be calculated with the help of the Biot-Savart law. It may be noted that the outward force on MBCN, i. e., the force acting in a direction in the plane of the circuit, and perpendicular to BC is all contributed by BC, the conductors MB and CN contribute nothing, as their lengths are parallel to this direction in which the force is required. (The force on a conductor in a magnetic field being at right-angles to the length of the conductor.) The outward force on MBCN is therefore the force due to MADN on BC, i. e., the sum of the forces on BC due to AM, AB, and DN.

Let us first calculate with the help of the Biot-Savart law the force on the conductor BC due to the conductor AM of the fixed portion MADN of the circuit-breaker. Take an element dx at Q in BC; $BQ = x$. The force on this element due to an element ds at P on AM ($PB = s$) is given by $\frac{i^2 ds}{R^2} \sin \theta$, where $PQ = R$ and the angle $\widehat{BPQ} = \theta$. But $R = s \sec \theta$, $x = s \tan \theta$, therefore we have

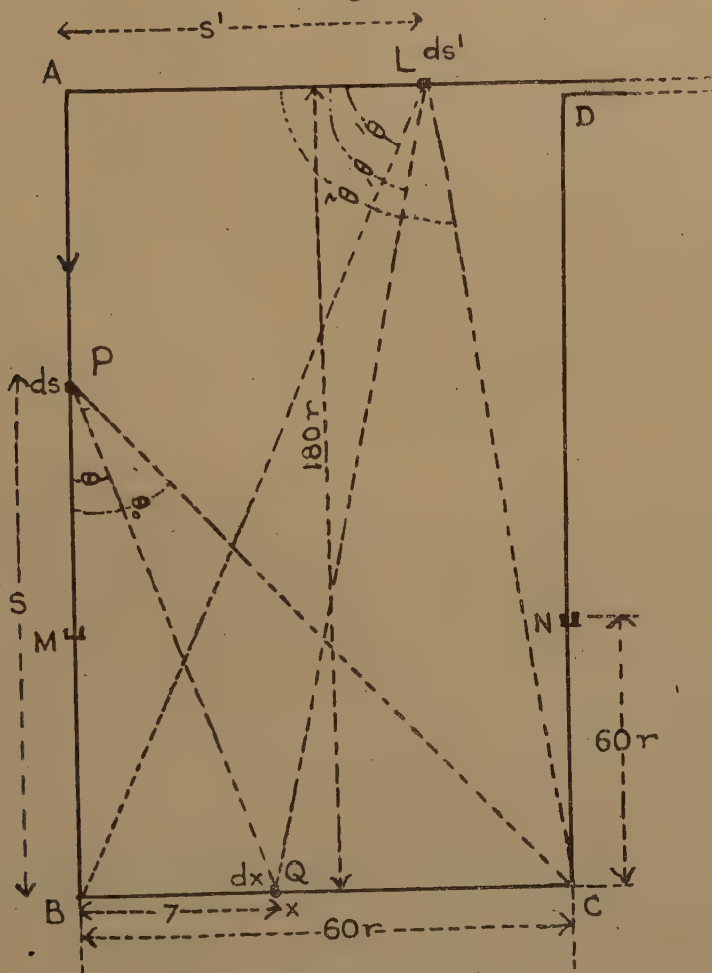
$$\text{the force on element } q = i^2 \frac{ds}{s} \sin \theta d\theta.$$

Hence the force on BC due to the element ds at P is

$$= i^2 \frac{ds}{s} \int_{\theta=0}^{\theta=\theta_0} \sin \theta d\theta,$$

$$= i^2 \frac{ds}{s} \frac{s}{\sqrt{s^2 + (60r)^2}} \Big|_0^{\theta_0} \Big],$$

Fig. 2.



Dunton's circuit-breaker.

and, therefore, the force on BC due to the conductor AM is

$$= i^2 \int_{s=60r}^{s=180r} \left[\frac{s}{\sqrt{s^2 + (60r)^2}} - 1 \right] \frac{ds}{s}.$$

$$= 0.162 i^2.$$

Similarly, the force on BC due to the conductor ND will be $0.162i^2$.

We have now to calculate the force on BC due to the conductor AD. The distance between dx and ds' , current elements in BC and AD respectively, is $AB \operatorname{cosec} \theta'$ or $180r \operatorname{cosec} \theta'$, and $x = s' - 180r \cot \theta'$. The force on the element Q due to ds' in AD is

$$= i^2 \frac{ds'}{(180r)} \sin \theta' d\theta',$$

or the force due to ds' on BC is

$$\begin{aligned} &= \frac{i^2 ds'}{180r} \int_{\theta'=\theta_1}^{\theta'=\theta_2} \sin \theta' d\theta', \\ &= \frac{i^2 ds'}{180r} \left[\frac{s'}{\sqrt{s'^2 + (180r)^2}} + \frac{(60r - s')}{\sqrt{(60r - s')^2 + (180r)^2}} \right]. \end{aligned}$$

Therefore, the force on BC due to AD is obtained by integrating the above expression for the limits $s'=0$ to $s'=60r$ which gives the value $0.108 i^2$.

Thus the total outward force on MBCN due to the circuit MADN is *

$$= (0.108 + 0.162 + 0.162)i^2 = 0.432i^2.$$

In the above calculation for the outward force on MBCN we assumed Newton's Law, *i. e.*, ignored the mutual forces between the conductors BC, BM, and CN composing MBCN. This assumption is erroneous, and we shall now find that the force when correctly calculated comes out much larger than $0.43i^2$. Considering MBCN, the conductor MB exerts on BC a force perpendicular to BC, and the conductor BC exerts on MB a force perpendicular to MB. These two forces are neither equal in magnitude nor opposite in direction, and *hence in calculating the outward force on MBCN these mutual forces between its components cannot be ignored for they do not cancel out*. In fact, to calculate the outward force on MBCN, *i. e.*, the force directed parallel to its length MB, we must take account of not only the force exerted by the circuit MADN, as we have done before, but also of the forces exerted by MB and CN on BC. The outward force on MBCN is therefore the force on BC exerted by the rest of the entire circuit, *viz.*, AB, AD, CD. It will be noticed that the outward force on MBCN now no longer depends upon the positions of M and N, whereas in the former calculation assuming Newton's law the value depended upon their positions.

* A similar calculation for the outward force on MADN due to MBCN gives

$$(0.031 + 0.108 + 0.031)i^2 = 0.170i^2.$$

Proceeding as before, the force on BC due to AB is found to be

$$=i^2 \int_{s=r}^{s=180r} \left[\frac{1}{\sqrt{s^2+(60r)^2}} - 1 \right] \frac{ds}{s} = 3.37i^2.$$

Similarly, the force on BC due to CD is $3.37i^2$, while the force due to AD has already been found to be $0.108i^2$.

Hence, the total outward force on MBCN (which is just the force on BC) is

$$=(3.37+0.108+3.37)i^2=6.9i^2.$$

The above calculations are only approximate, as we have not taken into account the current distribution inside the section of the conductors. We can obtain a more accurate result if we consider the change due to the displacement of BC in the magnetic flux linked with the circuit. In order to calculate this change we use the usual expression for the self-inductance of a rectangle.

The self-inductance L of a rectangular circuit of length n , breadth y , and radius of the conductor r is given by the following relation

$$L=(x+y)+4 \left\{ -2(x+y)+2\sqrt{x^2+y^2}+x \log \frac{2x}{r}+y \log \frac{2y}{r} \right. \\ \left. -x \log_e \frac{x+\sqrt{x^2+y^2}}{y} -y \log_e \frac{y+\sqrt{x^2+y^2}}{x} \right\}.$$

Differentiating with respect to x , we have

$$\frac{\partial L}{\partial x} = -3 + \frac{4x}{\sqrt{x^2+y^2}} + 4 \log_e \frac{2x}{r} - 4 \log_e \frac{x+\sqrt{x^2+y^2}}{y} + \frac{4y^2}{x\sqrt{x^2+y^2}}.$$

Substituting $x=180r$, $y=60r$, we obtain

$$\frac{\partial L}{\partial x} = 17.5.$$

The conductor BC, and hence MBCN, will, therefore, be subject to an outward force

$$\frac{1}{2}i^2 \frac{\partial L}{\partial x} = 8.7i^2.$$

Thus we see that the Biot-Savart law *when the inapplicability of Newton's law is recognized*, or the change of flux principle, gives for the outward force on MBCN a value about 20 times larger than that which is computed by applying the Biot-Savart law, but assuming (wrongly) the validity of Newton's law. Dunton found experimentally, as is to be expected, that the force is $8.7i^2$ and *not* $0.4i^2$. From this he concluded that the Biot-Savart law sometimes indicates only a small fraction of the actual force. However, as is evident, it is not the Biot-Savart law that is in error; but it leads to wrong results when the inapplicability of

Newton's third law is not recognized. Drysdale⁽³⁾ has already pointed out the mistake of Dunton in neglecting the forces on BC due to the conductors MB and CN when calculating the outward force on MBCN. The reason for including the effects of MB and CN or BC, which Drysdale has at least not explicitly stated, is the invalidity of Newton's law for mutual forces between current elements.

My thanks are due to Dr. D. S. Kothari for his suggesting the problem and the hospitality offered at the Physics Department, University of Delhi.

References.

- (1) Biggs, 'The Electromagnetic Field' (Oxford University Press), p. 26.
- (2) 'Nature,' vol. cxl. p. 245 (1937).
- (3) 'Nature,' vol. cxliii. p. 601 (1939).

Physics Department,
University of Delhi.

[The Editors do not hold themselves responsible for the views expressed by their correspondents.]

FIG. 1.



FIG. 2.



FIG. 3.



FIG. 4.

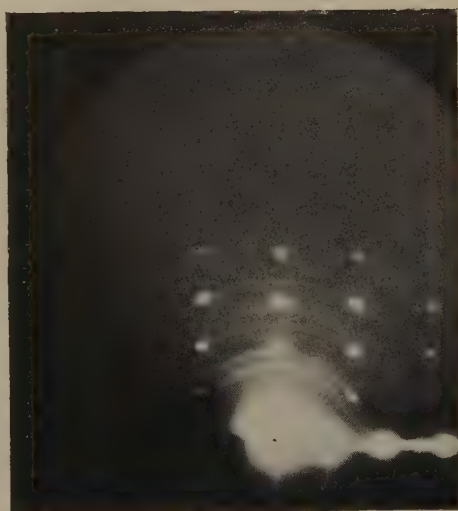


FIG. 5.



FIG. 8.



FIG. 9.



FIG. 11.



FIG. 12.



FIG. 14.

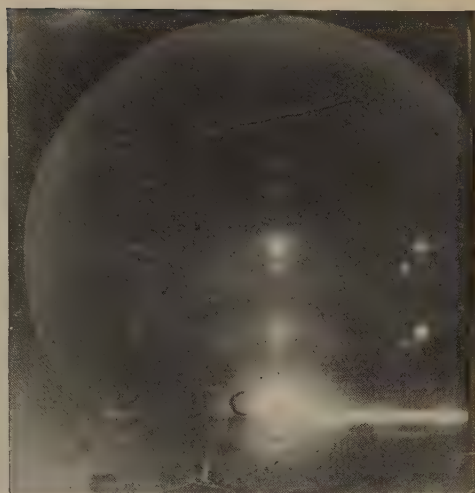
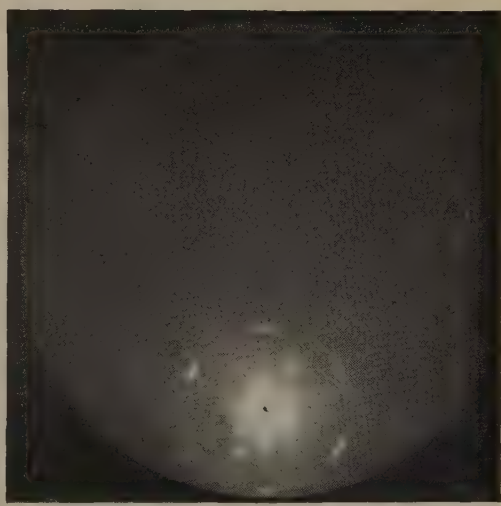


FIG. 15.



FIG. 17.



XIV. *Epistemology and Modern Physics.*

By HAROLD JEFFREYS *.

[Received May 9, 1941.]

1. It is well known that Sir Arthur Eddington claims to have proved that the whole of the fundamental laws of physics can be wholly derived from principles that he describes as epistemological, subjective and *a priori*. Laws not so derivable, he suggests, may occur in biology and psychology, but not in physics; and he admits that particular inferences cannot be derived in the way suggested, his arguments applying only to the most general laws, such as those of gravitation, electro-magnetism and quantum theory. If observation appears to confirm the laws, that is because the method of reduction used is such that the laws will apply to the results. He thus admits the usual view of logicians that observations are not predicted with certainty; but he maintains that the human wish for properties of conservation will always pick out such properties for study, and that he has shown that the numerical coefficients in them will necessarily have special values. An apparent difficulty in such a standpoint is that most of the laws obtained have, in fact, already been got (within various limits of accuracy) by the usual scientific procedure of combining observations; if they are *a priori* it is not easy to see how observations could ever have had anything to say about them. Eddington, however, appears to admit this also; the *a priori* character is introduced, not by the observations, but by the conventions introduced in reducing them. R. B. Braithwaite remarks † that he cannot close a very critical notice of Eddington's latest book without expressing admiration of his solitary heroism in defending such a desperate position. On the whole I agree, but I think that for a proper appreciation of the work it is necessary to examine first whether Eddington is using the words *epistemological*, *a priori*, *conventional* and *subjective* in their usual senses.

Theoretical science has three aspects: observation, imagination and analysis. Imagination is used in stating hypotheses for test; analysis in working out their consequences and comparing them with observation. In general several hypotheses can be stated in given circumstances, and the most probable is used for further prediction. I have developed in

* Communicated by the Author.

† 'Mind,' xlix. pp. 455-66 (1940).

my 'Theory of Probability' a formal system that enables this procedure to be carried out in full. The most important points to notice at the moment are that (1) if a hypothesis has led in the past to numerous predictions, all of which have been verified, the probability of future predictions from the hypothesis approaches certainty irrespectively of whether the hypothesis itself is true or not; (2) one of the fundamental rules is that any hypothesis capable of being stated sufficiently precisely to lead to verifiable predictions must be capable of attaining a high probability if a reasonable number of those predictions are successful. According to (1), a well-verified hypothesis may be treated as certain in future inferences; the probability of the inferences will be practically the same as if the more accurate inductive treatment was applied in detail. Thus deduction has a definite place in an inductive theory as an approximation valid for well-verified hypotheses, such as most of the laws of physics. This result amounts to saying that if another hypothesis is actually the correct one, it must nevertheless have led to indistinguishable inferences in the past and will presumably continue in most circumstances to give the same inferences; it is therefore convenient to regard alternatives not yet distinguishable in their consequences as together constituting a single hypothesis. Thus hypotheses are classified by their consequences rather than by their formal statement; if a modification should be found that gives different predictions (usually of some new type) it will be regarded as a separation of the original hypothesis into two rather than as something completely new. This is in accordance with usual scientific procedure. Chemists, for instance, are entitled to go on using the laws of chemical combination in their known domain without waiting until modern physics has explained why every known chemical compound should exist. The rule (2) is brought into the formal theory by the simplicity postulate. Since the verifiable content of the law is all that is relevant, a new phenomenon is expressed by the introduction of a new parameter, usually adjustable, into the probability of observations, given the hypothesis. To express absence of prejudice for or against the more complicated hypothesis I ordinarily take half the prior probability as concentrated in the simpler one*. The rule arranges

* Prof. E. A. Milne (Journ. Lond. Math. Soc. xv. p. 47 (1940)) suggests that a pre-incarnation of mine might have suggested, on the grounds of simplicity, that the sum of the angles of a triangle must be 180° . I cannot say for certain what he might have been silly enough to say, but the simplicity postulate would only say that the constancy of the sum of the angles had a finite prior probability until it was experimentally tested. The reason for the name "simplicity postulate" is that if the theory of probability did not contain it the observational evidence would always lead to the acceptance of the more complicated hypothesis. Since there is an element of chance in all scientific laws (Karl Pearson, 'Grammar of Science,' pp. 164-174), the postulate can be stated in

hypotheses in an order of complexity and, therefore, indicates an order of procedure ; but it admits any definitely stated hypothesis as capable of being accepted if the right kind of evidence exists. Like the rest of the theory, it is designed so that the fundamental principles will by themselves say nothing about observations or about scientific hypotheses, but so that they will provide a general method of comparing the two. These are, I think, the needs of epistemology. This may not be generally agreed, because it seems to me that some philosophers understand by the term a system that will support a scientifically unacceptable form of naïve realism. Eddington seems to occupy an intermediate position. His 'Mathematical Theory of Relativity' (referred to later as M. T. R.) starts with a four-dimensional interval as a primitive notion. His 'Relativity Theory of Protons and Electrons' (referred to as P. E.) introduces the general notion of a four-term relation as fundamental. In measurement of length, for instance, we are comparing relations between the ends of two objects. Now I should not regard the existence of lengths as epistemological. Mere non-identity of position is observable and, therefore, not epistemological. The beginnings of physics are in the recognition that differences of position are mainly systematic—in most cases if I see a blue patch to the right of a white one, I still see a blue one to the right when I look again, and so on. The difference in the amount of observational knowledge organized between this and the simplest measurement is obvious*. In speaking of a ds^2 , even for contemporaneous displacements, a further piece of observational information has entered, namely, that two bodies whose ends can be made to coincide when held in one orientation can still be so held in another orientation. If it were not for this fact it would be meaningless to speak of a length irrespective of the orientation. A great deal of what I should call physics has therefore been assumed in what Eddington calls epistemology. It is elementary and not controversial, but it is important that we should have a method of analysis that will allow us to test any hypothesis concerning observations whatever, and take nothing for granted, no matter how elementary. Consequently I think that his extension of the meaning of "epistemology" is highly undesirable.

the form "Variation is to be taken as random until there is positive evidence to the contrary."

Eddington (P. P. S. p. 4) says that he "can find no indication that the scientific researches of Pearson and Poincaré were in any way inspired or guided by their particular philosophical outlook." Some of Pearson's might have been better had they been more so guided, but the sentence as it stands is ungenerous to the inventor of χ^2 , the contingency table, and the best method of estimating the correlation coefficient.

* I have discussed the successive stages in 'Scientific Inference,' largely following Dr. N. R. Campbell's 'Physics: The Elements.'

I should similarly restrict "*a priori*" to propositions whose truth value is independent of experience ; therefore, in physics, it will include what I call epistemology, but not all of what Eddington does.

"Subjective" is a highly ambiguous term. Whitehead and Russell regard pure mathematics as both *a priori* and objective. If we regard everything as subjective that depends on the human observer or on human reason, then by definition the whole of science is subjective. Most philosophers, I think, would regard the external world, if any, as objective, and our sensations as subjective ; some would limit "objective" to what people can agree about and use "subjective" for matters where they differ. According to Eddington there is a subjective science, which includes the fundamental laws of physics, and an objective one, which includes psychology. It seems to me that we might well follow the good example of the systematic botanists and drop the terms as hopeless *nomina confusa*.

2. There is much confusion in the logical status of conventions. In their simplest form the attribution of a name is a convention, but it is useless unless there is such a thing, at least on some hypothesis that we are considering. In applications of elementary algebra it is conventional whether we call the price of oranges x and that of apples y , or conversely ; it is immaterial in the sense that when we translate back from symbolic to ordinary language we shall get the same prices of oranges and apples respectively, though not the same values of x and y . It is conventional whether we use rectangular or polar coordinates in specifying position. That is because the empirical properties of measured length are such that distances can be expressed in the form $(x-x')^2 + (y-y')^2 + (z-z')^2$; but if we like to take r, θ, ϕ as given functions of x, y, z , and work in terms of these, we shall always get the same relations between distances. The coordinate system expresses the choice of a language ; but the rules of the language must be such that the same thing is always said about the primary subject-matter. A convention is not something said, but something done ; but for completeness it requires the statement that it can be done, and this may be either an axiom, a theorem, or a result based on observation.

As an example of the illegitimate treatment of conventions we may consider the "methods of dimensions." The (largely empirical) properties of measures ensure that if we change a unit we alter all measures of that quantity in the same ratio. Now if we consider the equations of viscous motion of an incompressible fluid in the absence of bodily force,

$$\frac{du}{dt} = -\frac{1}{\rho} \frac{\partial p}{\partial x} + \nu \nabla^2 u, \quad \dots \dots \dots (1)$$

we see that, *keeping the units the same*, but multiplying all lengths in the

ratio λ , all times in the ratio τ , and all masses in the ratio μ , we get a similar system on a different scale. A solution for the original system will be adaptable to the new system provided that it still satisfies the equation, and for this it is sufficient that the three types of term shall be changed in the same ratio. Let p and ν be changed in the ratios ϖ and θ . Then the terms are changed in the ratios

$$\frac{\lambda}{\tau^2}, \quad \frac{\varpi \lambda^2}{\mu}, \quad \frac{\theta}{\lambda \tau},$$

which are equal provided that

$$\varpi = \frac{\mu}{\lambda^3} \frac{\lambda^2}{\tau^2}, \quad \theta = \frac{\lambda^2}{\tau}. \quad (2)$$

Thus if the linear dimensions of the system are denoted by L , representative pressures, velocities, densities and viscosities at corresponding points by P , U , ρ and ν , the solution for the first system can be directly adapted to the second provided that P is proportional to ρU^2 and ν to LU . If ν/LU is different the solution need not be thus adaptable; hence for geometrically similar systems

$$P \propto \rho U^2 f(LU/\nu). \quad (3)$$

This is the solution obtained by Stokes, by this method, that of dynamical similarity *. Now this is a perfectly valid argument, and gives what is wanted, a means of comparing stresses in different systems. It is to be noted that there are no changes of units, but the form of the equations of motion is a condition of the problem. In a compressible fluid, for instance, the equations are somewhat modified, and adaptation to the similar system is not possible unless, also, U/c is the same, where c is the velocity of sound. Then we should have to write

$$P \propto \rho U^2 f(LU/\nu, U/c). \quad (4)$$

One method of dimensions involves supposing the units changed, but the system itself kept the same. Thus the numerical measures are altered, but the equations will still hold provided the measures of pressure and viscosity are altered in specified ratios. But since we are still considering the same system, this tells us nothing new except how we should alter the numbers to be attached to given pressures and viscosities when we alter the units of length, time and mass. We have, in fact, made only a conventional change, and nothing but other conventional changes can arise in consequence, namely, how to alter derived units so as to keep the equations true.

* Rather curiously, the ratio LU/ν is usually known as the Reynolds number; though Stokes gave the result much earlier, and Reynolds did not even reduce his results to this form.

But the attempt is sometimes made in this and similar problems to use this method to derive the above formulæ for a different system. The argument is to derive the formulæ for the same system with different units, and then to use them for a different system with the same units; with suitable adjustment the numerical measures will be made the same and the formulæ are then transferred unaltered. This is a further step, and needs special justification. By itself it involves confusing a physical quantity with its numerical measure; it gives no meaning to the statement that 1 centimetre is different from 1 mile, and seems to say that one centimetre is different from 10 millimetres. It does, however, give the right answer. The reason is that the numerical assessments of the various terms are changed in the same ratio if the scale of the system is altered in one set of ratios, keeping the units the same, and if the system is kept unaltered and the units altered in the inverse ratios. If the equations hold in one case they will hold in the other. Consequently the mathematical expressions of the two methods are identical, though their meanings are entirely different. To compare different systems, using the same units, is not the same thing as to compare measures associated with the same system, using different units; and no amount of attention to the mathematical expression will remedy confusion about the nature of the subject-matter.

A still more ambitious method of dimensions is to ignore the form of the equations of motion altogether, and merely to note that the dimensional quantities associated with the system are L, U, P, ρ, ν, c . Since there are three fundamental dimensional quantities, length, time and mass, there must be three non-dimensional quantities associated with the above six dimensional ones; these can be taken to be $P/\rho U^2, LU/\nu, U/c$. If, then, there is any general rule connecting these it must have the form (4). But this begs the question; it assumes what is proved by the full set of equations, that no quantities other than the above six are relevant. In one respect it is more general, since the hypothesis might be right even if the form of the equations was different, but the equations themselves are the only guarantee of its truth. Actually the equations fail if the density is sufficiently large or sufficiently small, and other numerical ratios depending on the ratio of L to the atomic spacing will become relevant; and the result is then incorrect. But the deficiencies of the method as a proof need not blind us to its usefulness as a source of suggestions. It would suggest the form (4) even if the fundamental equations were not known, and then this form could be tested experimentally. It is actually verified over a considerable range of values, and since the function f has to be determined experimentally in most cases anyhow, little extra work is involved; the process of determining the function tests incidentally whether such a function exists.

I have dwelt at some length on this matter, in spite of the fact that it may be considered already settled*, because it illustrates how three methods, very similar in their mathematical statement, may rest on quite distinct bases; these should be explicitly stated, and if they are not a fundamentally confused argument will result.

Now there seems to be such a confusion in the general theory of relativity. This is partly suggested by Eddington's extreme casualness in the use of units. In M.T.R. he remarks that his coordinates are merely identification-numbers and have no reference to the unit of interval (p. 202); thus they may be ratios of intervals to some fixed standards of the same kind. But in finding the field of a point mass he uses r as a coordinate without mentioning this rule, and since his m/r is a pure number m would be a pure number if the rule was being applied. But he gives m for the sun as 1.5 kilometres and calls it the mass of the sun. He seems to be proud of this dimensional paradox, which appears in several of his works. What has happened is that the units have been chosen so that the velocity of light c and the constant of gravitation f will have numerical measure unity, and it has been supposed that they actually are the number 1. Hence length, mass and time are all treated as having the same dimensions. If it is remembered that their dimensions are different and that consequently f and c are not numbers, though their measures are numerically 1, we find that his m is really fm/c^2 , and its identification with a length is no longer paradoxical. He must, in fact, have used this form or something equivalent in order to get his value; but why not say so?

This kind of confusion had already occurred in electrical theory. We can write the force between two charges as

$$F = Eee'/r^2.$$

This is an empirical law, and e, e' are fundamental magnitudes in Campbell's sense. Electric charge is different in kind from length, time and mass, and must be assigned other dimensions. E is a constant of proportionality. Then we find easily that the dimensions of E are ml^3/e^2t^2 . The electrostatic unit is defined so that E will have measure 1; but it is not a number. If it is taken to be a number we shall infer the result that still appears in some text books, that e has dimensions $m^{1/2}l^{3/2}/t$. But this procedure is plainly illegitimate. For a similar argument would attribute the same dimensions to a magnetic pole, and a further treatment of the magnetic force due to a current defines the electromagnetic unit of current as of the dimensions of a magnetic potential. It follows that the ratio of the electromagnetic and electrostatic units of charge is a velocity. But this is complete nonsense, because the electrostatic charges

* Goldstein, 'Modern Developments in Fluid Dynamics,' pp. 99-101.

can be combined, and releasing then gives an electromagnetic current ; electrostatic and electromagnetic units are quantities of the same kind and must have the same dimensions. The difficulty has been habitually slurred over ; it plainly arises from the treatment of three different constants of proportionality as pure numbers when, in fact, they are all dimensional, and include two dimensions not expressible in terms of the three basic magnitudes. It is argued sometimes that if we changed the basic units we must make corresponding changes in the electrical and magnetic units ; but there is no such compulsion. The choice of the units involves conventions, and it is perfectly conceivable that we might continue to use the present ones, which are defined in terms of given standards and do *not* make the constants numerically exactly 1.

In P. E. Eddington appears to be again using dimensional coordinates, since he gives (p. 75)

$$ds^2 = -dx_1^2 - \dots$$

But now he has given standards of length and time, namely, h/mc and h/mc^2 , and the equations cannot be expected to be invariant for change of units. He is, in fact, using special units throughout. But he nowhere states their values in terms of the usual units, even though he must have the means of doing so, from his way of restoring suppressed constants when wanted.

Now the main feature of the tensor calculus, that equations must be transformable into any system of reference without change of form, is similar to the requirement that equations in modern hydrodynamics should be stated in a non-dimensional form. We must not, for instance, put the density of water equal to 1, even though we do not propose to consider any other fluid in an investigation. Using this rule, we can easily extend an experimental result to other fluids, or we can prove a theoretical one for all fluids as easily as for one. These are, however, the considerations involved in the method of dynamical similarity, not of either method of dimensions. With this rule the equations are invariant for changes of units, but that is incidental. Similarly the tensor property is a tautology. The constructive results of the general theory arise from the selection of *particular* tensors, and this involves physical hypotheses. Dynamical similarity, in fact, has a simple application. If we *assume* that the only fixed dimensional parameters in the laws are f and c , it will follow at once that if we have one possible solution, then another possible solution will be obtained by multiplying all lengths, times and masses by the same factor. Actually we know that quantum theory provides an absolute standard of length, so that the results cannot be expected to be valid for lengths comparable with h/mc . This has an analogue in the hydrodynamical case, since the results there would not hold for motion

about a solid whose size was comparable with the molecular spacing of the fluid. One hypothesis of the general relativity theory, then, is that there are no dimensional constants in the laws other than f , c and the cosmical constant λ . Given all three of these quantities we cannot rescale the system at all; if we vary λ all lengths will change with it. But then the measuring rods would alter in the same ratio, and epistemologically we should have the same system. A change of λ thus becomes unobservable, and we can see what Eddington means by his remark that the electron needs the local curvature of the universe to fix what size it should be. If the only dimensional fundamental constants are f and c , then the size of the electron is a fixed fraction of the radius of the universe. But the remark depends on this hypothesis. If there is another fundamental constant such as h/mc we could equally well say that the electron knows its own size and keeps it. The choice of the fundamental equations of general relativity is therefore not purely epistemological; a physical hypothesis has slipped in unnoticed by way of the hypothesis that only f , c and λ shall appear in them.

3. I have argued that Eddington's starting-point is not purely epistemological because it assumes the considerable amount of observational evidence that is needed before we can establish a measuring system at all. However, it may be agreed that even after assuming such a system it is surprising that so much should follow from its existence. Now Eddington starts in M. T. R. with the statement of an invariant interval ds between two events, and his further assumption is that everything can be deduced from information contained in ds^2 . But this involves accepting much more observational evidence and hypothesis. What justification is there for speaking of an interval between two events that differ in both time and position? To compare lengths we must inspect one pair of ends of two objects at the same time to see if they are in contact, look at the other pair similarly, and then look back at the first pair to see that they are still in contact. All the comparisons are contemporaneous, and an inductive inference is involved in supposing that the ends have not moved relatively between them. Similarly we can use a clock to measure the interval between two events where we are; we can compare clocks held together; but so far as I can see there is no possibility whatever of measuring an interval unless the events are coincident either in time or position relative to the observer. We can send a light signal over a measured distance and time its return; we thus arrive at a measured velocity of light to express the empirical fact that the distance and time of travel are proportional. We can then define a magnitude constructible from observed ones, namely,

$$ds^2 = c^2 dt^2 - dx^2 - dy^2 - dz^2,$$

where dt is the time interval on the observer's clock and dx , dy , dz are derived from distance measures made over the territory by the methods of mensuration. We can also verify the rectilinear propagation of light by studying the effects of obstacles. We can thus say that the times of transit of light behave as if ds so defined was zero for the path of a signal and $\int ds$ was stationary for small departures from such a path. Further, the rule that $\int ds$ is stationary for small variations of the path, even if ds is not 0, gives dx/dt , dy/dt , dz/dt constant, and thus gives a concise expression of Newton's first law—uniform motion in a straight line is expressible in terms of ds . But ds is introduced simply to give a convenient summary of these properties. The distance in space being defined in terms of length measures, equality of distance in different directions being defined by turning the same body round, it does not follow from the definitions that the time of travel of light there and back over the same distance is independent of the direction. This is proved experimentally for light by the Michelson-Morley experiment, and would not be true for signals, such as sound signals, travelling with any other velocity. It is this experiment that justifies the possibility of introducing an invariant interval ds combining variation of both position and time, and the properties of light are fundamental. It is therefore surprising that Eddington treats such an interval as fundamental and remarks (p. 19) :—
 “By good fortune there is an entity—light—which travels with the fundamental velocity. It would be a mistake to suppose that the existence of such an entity is responsible for the prominence accorded to the fundamental velocity c in our scheme ; but it is helpful in rendering it more directly accessible to experiment.” This is simply an inversion of the actual procedure. If there were no such entity there would be no justification for the introduction of ds and a fundamental velocity at all.

On the other hand, if we consider how the measures of two different observers will be related, supposing that $ds=0$ for either implies $ds'=0$ for the other, and that a particle moving so that $\delta \int ds=0$ for small variations of the path with reference to one will have $\delta \int ds'=0$ for small variations of the path with reference to the other, we can show that $ds'=k ds$, where k is a constant. This is true whether c is the same for both or not. It has actually been found that, using transported scales of length and standard clocks (ultimately compared with the standard metre and the rotation of the earth), we get the same c in widely separated regions. To fix k for comparisons between relatively moving observers a further comparison is needed, which could be obtained by means of a cathetometer, used to compare distances at right angles to the line of sight.

If it is found that rods so held when in relative motion, after being compared when held together, still agree in measured length, we can infer that $k=1$ and that ds defined in terms of measures is truly invariant for observers of the class considered. This comparison, however, can hardly be carried out as accurately as the others. Granting the result, however, the special theory of relativity can be developed. It will cover light propagation and the motion of unaccelerated particles. But it is necessary to emphasize the amount of experimental information that is summarized in the postulate of the existence of an invariant ds in four dimensions, once we have recognized that direct comparison of standards of length and time in relative motion is impossible. Even as it stands it may be held to contain elements that are not verified to the accuracy required. But the essence of any method is that it must be stated so that we can see *what the hypotheses are*. This is not possible when the existence of an invariant ds is postulated at the outset, for it involves what are intended to be measurable elements; until it is stated how these are to be compared we have attached no meaning to ds and ds' , and when it is stated it becomes an experimental question whether the alleged equality is true. We may regard the experimental way of comparing standards at different places as a definition; if so, $ds=ds'$ is a physical hypothesis (though possibly depending on observations that have not been made with sufficient accuracy). Or we can absorb k into ds' and then take the new terms as the definition of equal displacements in the four directions; then it becomes a physical question whether these agree with the experimental way of comparing standards. One rule may be a convention; it is impossible that both can be. Further, in any case the hypothesis that two observers in unaccelerated relative motion will regard the same particles as unaccelerated is physical, even though it must be regarded as only an idealized and possibly unattainable limit to what can actually be observed.

4. The general theory starts with the notion of an extended invariant ds ; and again it will follow that if $ds=0$ implies $ds'=0$ and $\delta \int ds=0$ implies $\delta \int ds'=0$, then $ds'=k ds$. The first postulate, as in the special theory, may be taken to represent the propagation of light, the second the statement that what appears to be a gravitational path for one observer will appear to be one for another. The second may be regarded as a natural physical generalization, particularly as another obvious possibility, that light is unaffected by gravity, is disposed of by the eclipse displacement of star images. Further, if we take

$$ds^2=c^2(1-2U/c^2) dt^2-dx^2-dy^2-dz^2 \quad \dots \quad (1)$$

we shall have

$$\begin{aligned}\int ds &= \int c \left\{ \left(1 - \frac{2U}{c^2} \right) - \frac{\dot{x}^2 + \dot{y}^2 + \dot{z}^2}{c^2} \right\}^{\frac{1}{2}} dt \\ &= [ct] - \frac{1}{c} \int (T + U) dt + O(c^{-3}), \quad \dots \quad (2)\end{aligned}$$

where T and U are the kinetic energy per unit mass and the gravitational potential. For variation of paths with the same termini, the condition $\delta \int ds = 0$ is therefore equivalent to Hamilton's principle except for terms of orders T/c^2 and U/c^2 in comparison with the terms in Lagrange's equations given by this principle. Consequently the suggested properties of ds^2 lead at once to results indistinguishable to this accuracy from those of Newtonian dynamics, when the coefficient of dt^2 in the special theory is modified by the addition of a simple multiple of the gravitational potential. The results of Newtonian dynamics can therefore be claimed, to this accuracy, as inductive support for the postulated existence of a ds with the assigned properties. But these properties still do not imply $k=1$, and therefore, the invariance of ds . This, as for the special theory, still seems to require the cathetometer test. There is a new difficulty here. For whereas in the special theory free particles travel in straight lines, in the general theory they do not. Body B after leaving body A will not travel the same path relatively to A as a light ray from A to B ; and the standard of rotation is not obvious. It seems possible that the curiously * named "parallel displacement" is relevant here; it should be possible to state a rule for the comparison of distant spheres by cathetometer measures based on common geodesic tangents. The present accounts of parallel displacement seem to me unsatisfactory. They depend on the idea of a "localized vector" in Newtonian dynamics. A localized vector, I think, is an unfortunate term. Two parallel forces are not equivalent unless their line of action is the same. Strictly a force requires the components of force *and the moments* to specify it. If forces act on different points of an elastic solid they are not equivalent even if they are in the same line, since if one of them is reversed the body will be strained. But it will not be strained if it is perfectly rigid, that is, if the velocities of elastic waves in it are infinite. Thus the equivalence of localized forces requires both that they act in the same line and that their points of application are on a perfectly rigid body. Otherwise there seems to be no meaning in equivalence of parallel forces so long as force is regarded as a vector needing the usual number of components to

* Curious because a vector carried round a small circuit by such displacement changes its direction, but not its length.

specify it. Similarly we need further analysis of the physical meaning of equality and parallelism of small displacements applied at different points. The standard works on relativity seem to take it for granted that there is such a meaning, but none that I have inspected gives the slightest indication of what that meaning is. In Euclidean mensuration we can, however, compare displacements at a distance by the cathetometer test when they are normal to the line of sight. The generalization of this result would be that geodesics initially parallel remain equidistant. (This, however, is false for geodesics on a sphere.) Similarly, there is a possible standard of direction perpendicular to the line of sight, namely, the direction of polarization of a light wave. It would be possible to define transverse displacements along a light ray as in the same direction when vibrations in their directions arrive at a given observer in the same plane of polarization. Thus light rays and their planes of polarization give a physical meaning to parallelism at any distance. Whether an extension is possible to points not on a null geodesic does not appear obvious. It remains to be seen how many of the results of general relativity would still hold; for this definition is in some respects more and in others less general than the usual one. But it seems that parallel displacement, instead of being a rather mysterious idea defined in terms of ds , is really a physical idea related to certain possible experiments that relativists have forgotten to mention, which are necessary before we can assert that a standard will give the same ds wherever it is.

The notion of a fundamental tensor, namely, one that depends only on the coefficients $g_{\mu\nu}$ and their derivatives, involves the same sort of considerations as were pointed out above with regard to the methods of dimensions. Once granted an assemblage of mutually transformable coordinate systems to identify events, the tensor property is a tautology like the statement that change of units does not alter the observed phenomena for a given system. In fact the equations of motion and the differential equations satisfied by the $g_{\mu\nu}$ might have any form whatever; they could always be transformed to any system of reference by the differential calculus, and the resulting equations would have the required generality. They would, presumably, be more complicated, and it may be held that the equations of general relativity are already quite complicated enough. But the obtaining of definite results is a matter of the choice of the *particular* tensors. This was originally done by considering the conditions that ds^2 in arbitrary coordinates should be transformable to the Galilean form by a suitable choice of coordinates. These conditions gave the relation $B_{\mu\nu\tau=0}^\sigma$ between the $g_{\mu\nu}$. Contracting this gave $G_{\mu\nu}$; this was taken to be zero outside matter and proportional to the energy tensor inside matter. Now the Galilean form is that of the special theory, the grounds for which have been discussed above; in this theory, all

particles move with uniform velocity, and we have an extension of Newton's first law. $B_{\mu\nu}^{\sigma}=0$ is then the condition for absence of a field. But nothing observable would be affected by a uniform acceleration of all particles and it is also the condition for uniformity of acceleration. But the latter condition in Newtonian dynamics is

$$\frac{\partial^2 U}{\partial x_{\mu} \partial x_{\nu}} = 0.$$

Contracting this tensor we get $\nabla^2 U = 0$, which is the equation satisfied by the Newtonian potential outside matter; and within matter $\nabla^2 U$ is proportional to the density. Hence the actual method of getting the equations of the general theory corresponds exactly to properties of the Newtonian field. The essential change is in the expression of the equations of motion by $\delta \int ds = 0$ instead of by $\delta \int (T + U) dt = 0$, the two agreeing in any case except in terms negligible when the velocities are small compared with that of light. The problem solved is to adapt Newtonian dynamics so as to make it consistent with a finite velocity of light. It has already been pointed out that the changes made are trivial compared with the correction for the time of travel of light in the comparison of Newtonian theory with observation. The need for a unified theory of light and mechanics was foreshadowed by Newton's own suggestion that light might be deflected by gravity, and reinforced by Laplace's idea that the most massive bodies in the universe might be invisible because the velocity of light would be less than the velocity of escape, a clear anticipation of the Einstein universe. The mathematical equipment was provided by Riemann. Why was it not till 1915 that the ideas were put together? I suggest that it was simply that the use of ds instead of Hamilton's principle did not occur to anybody until it was seen to be the best way of describing the Michelson-Morley result, and that the relevance of ds , which seems to Eddington an epistemological necessity, rests on this experiment and on the vast amount of observational evidence already existing in support of Euclidean mensuration and Newtonian dynamics. It may be recalled at this point that to make a small alteration in an inductive theory is not a contradiction; the possibility of making such an alteration is essential to the inductive method.

One peculiarity of Einstein's method is that it does not postulate the equivalence of observers. The whole argument is carried out in one coordinate system. The equations are necessarily in tensor form and, therefore, the method of transforming to other observers is mathematically straightforward; the physical difficulty of comparing different observers' standards remains. So far as one observer is concerned the equation $ds' = k ds$ does not arise. But this does not dispose of k . For one observer

must still consider intervals measured in terms of a transported standard, and he can still observe that standard, in principle, however far away it is, thus comparing the same standard at different distances from himself by the cathetometer test. It is then a physical hypothesis that ds for intervals of the transported standard will be the same for one observer wherever it is. This is evidently the hypothesis used in inferring the spectral shift, the detection of which may be considered as a confirmation of it. The cathetometer test would compare transverse space-like intervals at different distances, while the spectral shift compares wavelengths in the direction of sight, but the principle is the same. But also, recognizing the physical character of the hypothesis, we may admit the logical possibility that differences of ds may arise in such displacements. If they do we appear to have something analogous to the rather mysterious gauge-transformation of Weyl *. What this argument does is to eliminate the second observer and leave the residuum of physical hypothesis that is verifiable by one observer. The results can then be extended to all observers. We have not used the condition of equivalence of observers, because that must be a tautology for some way of stating the laws; what the method does incidentally is to discover what way that is.

The argument as discussed above deals only with consequences of the postulates concerning ds and therefore cannot express any phenomena not describable in terms of ds . For some such reason Eddington regards the theory as a geometry; but since the geometry has been defined so as to include the whole of mensuration and gravitational theory, he is using the word in a sense very far from its usual pure mathematical one. He regards gravitation as a geometrical phenomenon; naturally it is, in his sense, because ds and its properties have been assigned so as to include it. Whether words should be used so far from their usual meanings when other suitable words for the new meanings exist is another matter. Equally naturally, in his sense, geometry does not include quantum or electrical phenomena; for these do not depend wholly on mensuration and gravitation.

5. Eddington's contributions to quantum theory have aroused much criticism from several sides. That is doubtless in part due to their extreme importance if they are correct. It is certainly a tremendous advance if the fine-structure constant, the ratio of the masses of the proton and electron, the constant of gravitation and the number of particles in the universe can be calculated from other data with more accuracy than direct experiment gives—indeed, three of them with complete accuracy. The bare possibility of such determinations had not,

* It becomes still more mysterious in P. E., since Eddington now finds that it must be imaginary.

I think, been considered by any previous writer. On the other hand, few seem to find his arguments convincing. My personal reaction is that the astonishing agreement with experiment suggests that there must be something in the argument, but as I have several times had to deplore that attitude in other people I am reluctant to accept it myself. I am very conscious of my own insufficiency for a criticism of work that is far from my special field, but think that, at any rate on the epistemological side, I can offer something useful, particularly as in some respects Eddington seems to me to be less unsatisfactory than some of his chief physical opponents. The situation is very different from that with respect to M. T. R., where I think that I can disentangle what the hypotheses really are, and the results follow from them. In P. E. I think that nobody knows what they are or whether the conclusions follow from them.

The fundamental idea in P. E. is that of E-numbers, which are linear functions of certain symbols satisfying the relations $E_\mu^2 = -1$; they are more general than the usual algebraic square roots of -1 , there being, in fact, sixteen of them. Fifteen of them form three pentads, the members of which satisfy the anti-commutation relations

$$E_\mu E_\nu = -E_\nu E_\mu.$$

The other is the usual algebraic root. Thus at the start there is a considerable extension of ordinary algebra. There is no mathematical difficulty here, such an algebra being perfectly consistent. The question is, why so complicated an algebra is necessary to describe physical facts. Observed quantities are always real; quantum theory is said to deal with probabilities, which are essentially real and positive. Why should it be necessary to introduce even the ordinary square root of -1 ? The answer, presumably, is that it is not, but it may be a great convenience. For instance, in two-dimensional potential theory we know that the electric force is the gradient of a function V satisfying

$$\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} = 0,$$

the general solution of which is

$$V = f(x+iy) + g(x-iy),$$

and the fact that V is real shows that the functions $f(x+iy)$ and $g(x-iy)$ must be conjugate complexes. Thus we can write

$$x+iy=z, \quad V=2\text{R}f(z),$$

R denoting the real part. But then if we write also

$$2f(z)=V+iU$$

we find that the difference of the values of U at two points is the total flux of electric force across any curve joining them, and in particular if the points are on the surface of a solid this difference leads at once to the charge lying between them. This is a case where the introduction of complex quantities is demonstrably a great convenience in the handling of real ones. But in accounts of quantum theory we are presented on the first page with at least $\sqrt{-1}$, and usually with matrices or numbers not satisfying the commutative rule of multiplication. These cannot be necessary. They may be convenient, but if so it should be shown that the rules satisfied by the probabilities are such as to make them convenient. This is never done, and Eddington is the only author who has even attempted it. The rules satisfied by the probabilities are never even stated, and actual assessments of probabilities emerge from the analysis without any indication of what the previous steps mean, apart from the fact that the results of the complex analysis must be combined in such a way that the answer turns out real when verifiable results are to be produced. The situation is that the physical meaning of the postulates is wholly obscure, and it is completely mysterious what is the essential property of the answers they lead to: there must be some such property replacing the rules of classical mechanics, and it should be capable of some much more direct statement*.

Now Eddington does try to explain the purpose of E-numbers. He points out that, for instance, a displacement requires for its existence the existence of the two relata; we may write 1 if either relatum exists and 0 if it does not, so that the existence of either relatum may be denoted by an operator J whose eigenvalues are 0 and 1. That of the displacement will then involve the product of two such operators J_r, J_s . In either case $J^2=J$ and the operator is called idempotent. Now a matrix must be 2×2 to have two eigenvalues, 4×4 to have two independent pairs; but all 4×4 matrices can be expressed in terms of E-numbers, and the introduction of the latter appears to be explained. But the argument clearly does not show that the E-number analysis is equivalent to the statement of existence of two particles, though the latter is all that appears to have been assumed. For the latter is expressed in terms of four possible pairs of numbers, all real and unique; the result is expressed in terms of sixteen matrices, many of which are imaginary. At the most four numbers would say all that can be said on the hypothesis, but sixteen E-symbols are introduced. If they are needed they plainly express much more than the hypothesis, and, in fact, do so in Eddington's analysis since he expresses in terms of them not only the components of the

* Mr. J. H. Moyal and I, in unpublished papers, have found partial explanations.

displacement but the spin and the electric charge. He has said, in fact, much more than his hypotheses warrant by introducing eigenvalues and matrices at all. If the elements of a matrix are a_{ij} and x_j is a vector whose components are equal in number to the columns of the matrix, and if

$$\sum_j a_{ij} x_j = \lambda x_i,$$

where λ is a number, then λ is called an eigenvalue and x_j an eigenfunction. But for vectors other than eigenfunctions the operation is still mathematically significant. Thus, by speaking of eigenvalues Eddington has implicitly assumed the existence of a four-component quantity to operate on. This appears to be a wave-vector; but the immediate point is that when we have solved a quadratic equation the roots may be the end of the matter; there is no necessity for them to be eigenvalues of something more complicated. The proof of the four-dimensionality of space-time from the mere notion of the existence of particles therefore fails, and the justification of introducing the E's is no better than the fact that they enable us to write

$$ds^2 = -\sum dx_\mu^2 = \left(\sum_{\mu=1}^4 E_\mu x_\mu \right)^2,$$

the four-dimensionality of space-time being already assumed. The first four E_μ will only deal with the special theory of relativity: the rest are needed to deal with gravitational, electromagnetic and quantum phenomena. Thus the deductive argument from epistemology fails at the start, and information from other sources has been put into the structure of the theory. What the argument proves, if the theory holds, is that the E-number language is capable of expressing all the laws of physics, but the ground for believing that must be empirical*. Here we come again on the principle that arises in the second method of dimensions, where, if we knew that no parameters other than those stated were relevant, we could assert that there is a functional relation between them; but to assume this is a physical hypothesis. Similarly, in general relativity, to insist that all laws must be derived from the fundamental tensors assumes that nothing but the $g_{\mu\nu}$ are relevant to motion, and this is just what needs

* Braithwaite's analysis in the review quoted above merits attention; I have deliberately made my remarks overlap his as little as possible. He is wrong, I think, in claiming to deduce a mathematical contradiction from Eddington's argument. Eddington introduces an operator $2J-1$ with eigenvalues ± 1 , but Braithwaite argues that on Eddington's postulates it would always be $+1$. But the transformation is merely a change of language. If J can be 0 or 1, $2J-1$ can certainly be 1 or -1 . Braithwaite has supplied the postulate that $J_r + J_s$ is a J , but there is no need to do so, and if $J=0$, $2J-1$ is not a J in the original language.

observational verification. The process is the universal one of the inductive method, that we do not introduce new complications until there is positive evidence that they are needed. Eddington claims that his process is "mainly deductive," but this would apply to any good inductive theory, since in such a theory the hypotheses must be stated and their consequences worked out deductively. I think that the claim that induction is not used is disposed of, but it remains possible that the theory is deductive in the sense that, for instance, Love's 'Mathematical Theory of Elasticity' is a deductive work. The latter starts with Hooke's law, extends it to three dimensions, and the rest of the work consists of solving differential equations for various types of boundary condition. The experimental evidence for Hooke's law is not part of the subject matter, but it would hardly be denied that without this evidence and the possibility of applying experimental checks to the results the work would have little interest. Now Love's book satisfies the first requirement of both deductive and inductive theories, namely, that the postulates are explicitly stated. No modern quantum theory seems to satisfy this condition; at any rate the postulates are stated in such forms that it is impossible to say what they mean, and in some cases they mean something that is certainly false. On the first point I refer again to the immediate introduction of imaginaries, on which quantum theory seems to be in the state that two-dimensional potential theory would be in if we did not know Laplace's equation and treated the potential as the real part of a function of the complex variable. On the second, it is sometimes said that a joint probability distribution for the coordinates and momenta is intrinsically meaningless—not merely what is said by the uncertainty principle, that the product of their uncertainties cannot vanish*. But quantum theory is a modification of classical mechanics, in which the motion of bodies is predicted, the data being the initial coordinates and velocities. If their simultaneous probability distribution was meaningless, so would the predictions be; but they not only can be made, but are verified. A highly distinguished physicist told me recently that there is no real contradiction here, but that these problems are in the domain of classical mechanics, and quantum theory deals with an entirely separate domain, that of steady states. I do not think that the matter is so simple, however. I doubt whether a strictly steady state ever occurs. A heated wire does not wait till it has reached a steady temperature and then begin to radiate, and such a separation is impossible.

* A. J. Ayer ('Language, Truth and Logic,' p. 214) points out that the criterion that a statement, to be significant, must be capable in some circumstances of being made with certainty, would make every empirical statement nonsensical.

A complete quantum theory must be able to deal with varying states, and its full statement will give the whole of classical mechanics as an approximation, not merely the part of it that deals with steady states. But to deny the possibility of a joint probability distribution of the coordinates and momenta slams the door against any such attempt*.

The trouble seems to start with the idempotent operator. We notice first that there is really an idempotent operator in probability theory. If, as usual, $P(p/q)$ denotes the probability of p on data q , we can introduce an operator J depending on r, q such that

$$JP(p|q) = P(pr|q).$$

for all p, q . Then

$$J^2P(p|q) = JP(pr|q) = P(prr|q) = P(pr|q) = JP(p|q),$$

and $J^2 = J$ is an instance of the axiom that equivalent propositions have the same probability on the same data. Now we may compare this with the argument in quantum theory. Temple† remarks that a certain type of experiment impresses its character on the system. If this operation in his notation is denoted by P , then if the same system is immediately subjected again to the same process it will undergo no further change, and $P^2 = P$. But in practice there will always be a finite interval of time between two experiments; $P^2 = P$ can have application only if they are done simultaneously, and thereby become the same experiment. The use of $P^2 = P$ in practice means that the quantity observed has not changed between the two experiments. In the probability analogue we should have to suppose an additional datum q' to express the fact that some time has elapsed, but then we must introduce a new J to express the operation of including r in the proposal on data qq' . Call this J' . In

* On the other hand, C. G. Darwin (Proc. Roy. Soc. A, cxvii. p. 273 (1927)) gives a probability distribution for the position of an electron under constant electric force which could be written, changing his parameter σ to $\sigma\sqrt{2}$ to agree with the standard error as usually defined,

$$P(dx | Ht) \propto \exp \left[-\frac{(x-x_0-vt-\frac{1}{2}gt^2)^2}{2\{\sigma^2+(ht/4\pi m\sigma)^2\}} \right] dx.$$

This is the same as

$$x = (x_0 \pm \sigma) + t \left(v \pm \frac{h}{4\pi m\sigma} \right) + \frac{1}{2}gt^2,$$

the uncertainties being taken as independent. The probability is exactly the same as in classical mechanics, the initial position and momentum having standard errors σ and $h/4\pi\sigma$, and the errors being uncorrelated. For non-uniform fields, especially magnetic fields, there would be correlations, because the accelerations would also be affected by errors in the position and velocity.

† Proc. Roy. Soc. A, cxxxviii. p. 481 (1932).

general $J \neq J'$, but $J = J'$ if q' is irrelevant to r . The physical assumption in writing $P^2 = P$ is that any changes in the system between the experiments are irrelevant to the observable; in other words, that the observable is constant. This is evidently a severe restriction on the experiments permitted. In the first place, since in the first trial the state considered is that after the experiment, for symmetry the same must hold after the second. But strictly the second experiment must disturb the system, and it seems that the statement contradicts the uncertainty principle, unless indeed the experiment is confined to simple observation of the steady states of systems, the observer himself being a part. In the second place observables are limited to quantities that remain constant. In the solar system we should be limited to the masses, the linear and angular momenta and the energy. These are far from enough to predict the motion, but classical mechanics can do it. Passages in Eddington's writings suggest that he does not accept this limitation, but he does argue that it is an epistemological fact, or, alternatively, a peculiarity of the human mind, that we select those things that are constant as being of special interest. Now the argument of the idempotent operator would apply to constants, but in this sense is it epistemological? It is essential to a law that when the parameters in it are given, the unexplained variation is random in the sense that the errors in previous observations are irrelevant to that in a new one. But there is nothing to say that other laws involving additional parameters cannot be required for other observations. Inductive method is capable not only of showing that there is evidence for particular laws, but also of altering them if need be. Then it appears that Eddington has started by adopting the hypothesis that the fundamental laws have certain forms, involving three fundamental constants (say m_e , h , and c), and found that these require other constants to have certain definite values in terms of them. If these are right, there is inductive evidence that he has chosen the correct fundamental laws; if they are wrong there would be evidence that modification of the laws is needed. That is to say, taking the work at its face value, there is no evidence that it is anything but the development of a scientific hypothesis of the usual type, to be judged according to the agreement of its consequences with observation, and therefore inductively. Eddington argues that we should not be satisfied till we had reduced all fundamental quantities to constants. Perhaps that is so, but it begs the question to say that we have now reached a stage when we should be satisfied, or that the only fundamental quantities are those that we have now reduced to constants.

Eddington gives an apparent refutation of the usual view that *a priori* scientific knowledge is impossible*. "It has come to be accepted as

* 'Philosophy of Physical Science,' pp. 25-26 (referred to later as P. P. S.).

a scientific principle that we can have no *a priori* knowledge of the universe. Agreed ; provided that by "universe" we here mean "objective universe," as was undoubtedly intended when the principle was framed. But, as applied to a universe defined as the theme of physical knowledge rather than by its intrinsic characteristics, the principle cancels itself. If we cannot have *a priori* knowledge of the universe, we cannot have *a priori* knowledge that it is objective : and therefore we cannot have *a priori* knowledge that we cannot have *a priori* knowledge of it." The argument misses the point. There is no need to speak of the universe. In an inductive system it is entirely arbitrary whether we regard the world as having an independent existence or as a mental construct to coordinate our sensations, though libraries have been filled with bad arguments on both sides. Eddington propounds an unanswerable question in place of the relevant one : Can we predict sensations *a priori* ? He admits that we cannot so predict individual sensations, but says that we can predict the regularities connecting them, since these are imposed by the method of analysis. But to establish that he would need (1) to show that his method is purely *a priori* and that he has not been introducing physical hypotheses at every turn without noticing them ; (2) to explain just how the analysis of experiments designed to test some fundamental law imposes these regularities ; on the face of it the method of analysing observations is equally capable of showing that the suggested regularity does not exist.

6. Since modern quantum theory deals mainly with probabilities, and in particular Eddington's work is said to do so, it is strange that he does not give a single actual probability distribution, with the exception of uniform ones intended to express complete ignorance. He does emphasize that "probability is an attribute of our knowledge of an event ; it does not belong to the event itself, which must certainly either occur or not occur" (P. P. S. p. 50). If this was maintained I should be glad to welcome a distinguished convert to the view that I have been maintaining since 1919. I welcome also his statement on the next page : "The statement often made, that the electron is not a particle but a wave, is misleading. The wave represents our knowledge of the electron The terminology of quantum theory is now in such utter confusion that it is well-nigh impossible to make clear statements in it." But, unfortunately, he also says (p. 95) : "In physical science probability is essentially a statistical conception, that is to say, it is *defined* as a frequency in a class of events." One definition or the other might be taken as fundamental ; they cannot both be. The relation between them needs much further analysis, which I have tried to supply elsewhere. The first definition is, I consider, always the fundamental one. Estimation

of a probability from an observed frequency always involves the assumption of an empirical law, which, however, is always a law relating probabilities in the first sense. Again, in P. P. S. p. 92, Eddington considers "two similar bags A and B : A contains two white balls and one red one, B contains two red balls and one white. We draw a ball from one of the bags and find that it is white. It can be deduced that the chances are 2 to 1 that the bag is A." This is a simple application of the principle of inverse probability, which is a theorem in my system. But Eddington has on several occasions attacked this principle, in particular in one paper written after this passage*. On a previous occasion I pointed out† that in an attack on it‡ he had made all the hypotheses needed to prove it. Again, in P. E. p. 281, he writes: "We represent the actual probability as the product of two factors: (1) the initial probability, or "basis of statistics," and (2) a modifying factor which incorporates any special information supplied." If this is not a statement of the principle of inverse probability, what is it?

However, there is a fundamental verbal confusion in the expression "*a priori* probability," which has been used in three different senses in probability theory; two of them persist in Eddington's "Initial probability." I prefer to use "prior probability" for the probability of the hypothesis given the previous information; this may be only the minimum of such information needed to give the hypothesis a meaning. In the passage just quoted Eddington is using "initial probability" in this sense. But some writers speak of the *a priori* probability of getting seven heads in twelve tosses of a coin as 2^{-12} $^{12}C_7$. This is the probability of an observational result given a hypothesis, and Fisher's term "likelihood" is available for it. Now the so-called *a priori* probability of statistical mechanics is an approximation to the likelihood, not a prior probability. Assuming classical mechanics, we can show that a given element $dq dp$ in the coordinate-momentum space keeps its volume unchanged throughout the motion, but its shape is drastically altered. Starting with any information about the initial state of a system, not in general a uniform distribution with regard to the q 's and p 's, the distribution of the probability at a later instant is given by $f(q_0 p_0) dq dp$; q and p now referring to the later instant and $q_0 p_0$ to the corresponding initial values. But in actual observing conditions, on account of the distortion of the elements $dp dq$ during the motion, many different filaments corresponding to different ranges of q_0 and p_0 , and therefore to different values of $f(q_0, p_0)$, will cross a fixed element $dq dp$. Hence the probability distribu-

* M. N. R. A. S. c. p. 360 (1940).

† Phil. Mag. xxii. pp. 337-359 (1936).

‡ Proc. Phys. Soc. xlv. pp. 271-82 (1933).

tion with regard to $dq dp$ will tend to uniformity. The only proviso is that since the first integrals for the whole system will not change, values of q_0, p_0 corresponding to different values of these integrals will not mix, and the probability distribution for them given initially will persist; but this is dealt with by a known theorem *, which allows us to consider their values as data. The probability that any function of the variables after a long time will lie between definite limits is then found by integrating $dq dp$ over all values that make the function lie within those limits. This method leads to Maxwell's main results. A fuller statement of it also indicates clearly in what conditions those results might fail †. The usual presentations are utterly fallacious. But though the final result means that we shall know nothing about the coordinates and momenta after a long time, except whatever we knew initially about the first integrals, it is not a prior probability but a probability distribution of the variables given a certain hypothesis, namely, the laws of classical mechanics and any other information about the initial values, and therefore is a likelihood.

Presumably if the full laws of quantum mechanics were extended to cover unsteady states they would lead to the corresponding quantum form of the distribution function, which will still be a likelihood. But here there is a confusion, for which, strange to say, the late Lord Rayleigh is responsible. He applied the argument from equipartition to show that energy in a system would ultimately concentrate in the most rapid normal modes of oscillation. But Rayleigh knew better than most people that if a system has normal modes, energy initially in a normal mode will stay in that mode. If energy is not originally uniformly distributed between the modes it will never become so. This, in fact, is a case where the above argument fails. It shows only that the uniform distribution of probability will be approached subject to those quantities remaining constant that are required to be constant by the classical equations, and these quantities include the energies in the normal modes separately. It does not seem to have been suspected that the failure of Rayleigh's law of radiation to agree with experiment might be the result of a mistake in the theory. A fuller analysis suggests that if account is taken of the second-order terms in the equations, the separation into normal modes will only be approximate, and there will be a slow interchange of energy, Fourier analysis giving a concentration into frequencies of the order of those of the lattice vibrations of the enclosure. Detailed analysis is hardly needed because we now know otherwise that the classical equations of motion need modification, but the matter is worth

* 'Theory of Probability,' Theorem 11, p. 40.

† Jeffreys, Proc. Roy. Soc. A, clx. pp. 325-48 (1937).

mentioning because the apparent prediction from Rayleigh's argument that steady states including radiation are impossible on classical theory is still often made one of the foundations of quantum theory.

Eddington remarks in *P. E.* that the normal way of combining probabilities is to multiply them, but sometimes they may be added. This is a really remarkable statement. Probability theory contains both an addition rule for the probability that one of two exclusive alternatives is true, and a product rule for the probability that two non-exclusive alternatives are both true. The former is the more fundamental; in my treatment it is a convention, which introduces numbers to express probabilities in the first place. The second, which depends on the first, is a theorem in most cases; it is not clear whether these cases include all those where we need to use the rule or whether it must sometimes be extended axiomatically. Eddington seems to find great difficulty in these two elementary principles, for instance on p. 154, where he seems surprised that the probability that two particles are both in unit volume is not the same as the probability that at least one of them is. Nevertheless he does arrive at a correct epistemological idea of interaction, namely, that the probability-distribution for the position of one particle by itself is different from the corresponding distribution given that another is present. Similarly the physical constants, in the last resort when the relations of hypotheses to the data have been thoroughly analysed, will be parameters in probability-distributions.

Now Eddington gives a prominent place to the particle of mass m_0 , which is not a known particle, but simply either an electron or a proton, we do not know which. The difference of the masses of the electron and proton is simply an expression of the fact that they are disturbed by different amounts by equal electric charges in their neighbourhood. He proceeds to consider a charged particle and an unspecified particle, without interaction, and to show that information for the two can be expressed either (1) as the product of their wave vectors or (2) as a double wave tensor involving both. The two wave vectors and the tensor require respectively 10, 1 and 136 dimensions to represent them. Eddington argues that according to the method adopted we shall adjust the masses of the particles, which are eigenvalues of the operator $i\partial/\partial t$. This leads to the equation

$$10m^2 - 136mm_0 + m_0^2 = 0,$$

the roots of which will be the masses of the proton and electron. Now how does the neutral particle manage to have only one degree of freedom, namely, the time? One might expect it still to have the three position coordinates. Again, why should a particle so defined have a wave function? The probability that it will be in a given element of volume, given that

it is either an electron or a proton, we do not know which, is simply the mean of the probabilities that an electron and a proton respectively will be there. But what reason is there for supposing that such a mean is factorizable and thereby is capable of being treated by the wave-vector methods? Eddington provides none. Again, since he speaks of the alternative representations as an alteration of the "basis of statistics," which he elsewhere identifies with his initial probability distribution, he may really be considering a statistical problem of many particles of specified charge and many of unspecified charge that have been left for a long time, so that their probability distributions have become uniform. But for this to happen there must not be enough unbalanced total charge for the mutual repulsions to overcome gravity and prevent the attainment of a steady state, and the analysis into a single particle of specified sign and one of unspecified sign breaks down.

Eddington's result that the ratio

$$hc/2\pi e^2 = 137$$

depends on an interesting idea, namely, that as the exclusion principle and the Coulomb force both become important in the same circumstances they may both be aspects of the same phenomenon. This may well be right, but the phenomenon cannot be what he says. He starts from the fundamental position of not knowing even the distinction between a proton and an electron. In such a case there is no meaning in the distinction between the statements "the proton is at A and the electron at B" and "the proton is at B and the electron at A." Eddington introduces an extended wave function, giving probability $\cos^2 \frac{1}{2}\chi$ to the one proposition and $\sin^2 \frac{1}{2}\chi$ to the other, and treats χ as an ignorable dynamical coordinate. But if there is no distinction between the two propositions it is meaningless to attach separate probabilities to them. We can only consider their disjunction, "one particle is at A and the other at B." The introduction of χ assumes at least that there is a significant difference between the propositions. Now a curious fact about the work is that χ appears in the resulting equations in the same way as the time does for a single particle; time itself does not appear explicitly because both particles are supposed referred to the same standard time fixed by their centre of mass. But the introduction of a special coordinate to express the probability of a mistake is open to the same sort of objection as was made above to the necessity of introducing 16 E-symbols to express a double-existence operator capable of only four values, three of them 0. We may make a mistake of identification, but if we do it will be a complete mistake; there is no need for a continuous variable to express it. Again, Eddington argues that at the stage of knowledge that he is considering the difference of mass between the proton and electron is not a

means of identification, because mass is defined by the operator $i\partial/\partial t$, and cannot be determined until we have some means of identifying the same particle at different times. I accept the last statement for other reasons. But if we have two particles originally widely separated, so that they can be observed independently, it would be perverse not to identify them for different times of observation by choosing the smooth paths, and this criterion suffices. The trouble begins when they are really close, as in a hydrogen atom, but Eddington's argument would apply to any distance.

A clue to what has happened is provided by a footnote about a double star with a large orbital velocity in the line of sight. The observer may identify the components by colour, but if he does the velocities will not obey the laws of dynamics. I should reply that if he is able to observe the colours he will also be able to observe the transverse velocities and the changes of colour within a half-period. The latter would cast doubt on colour as a means of identification, and continuity of transverse velocity would support the contrary identification. Nothing so advanced as the laws of dynamics is needed, the mere fact that the changes are continuous with one method of identification and discontinuous with the other being sufficient. But if we were to persist in using colour, the identifications would be associated with a reversal of the transverse velocity every half-period. Thus the error of identification would be periodic *in time*; and the time in question would be that corresponding to the centre of mass of the system, or of the external particle in Eddington's language. This has disappeared from his equations, but I suggest that it has reappeared disguised as a permutation coordinate. Why, in fact, should it disappear? The apparent reason is that it is the same for both particles and therefore would not appear as a dependent variable in the equations for the internal particle, which concern differences of coordinates. But that is no reason why it should not appear as an independent variable. Our possibilities of observing microstructure would not permit us to determine the relative orientation of the particles in a hydrogen atom, but it is still a significant statement that it is periodic in time.

7. I do not think that there has been any explicit mention of a difficulty in any theory that attempts to combine relativity, quantum and probability theories. The problem of prediction is to infer a person's future observations from his past ones, including usually reports from other people. But relativity theory is fundamentally concerned with the finite velocity of light. Consequently observational information can deal only with events up to time $t-r/c$ for the special theory, not up to t . Another observer's information will overlap that of the first, but they

cannot be identical, nor can either be derived from the other by any coordinate transformation. We might transform future into past by a relativity transformation, but not ignorance into knowledge. But solutions of the relativistic wave equation are applied right up to time t for all places, and the data for this cannot exist in general. It may, however, be possible to carry out the extrapolation in some cases, notably for an isolated system such as a hydrogen atom, subject to its not having absorbed or emitted a photon since it was last observed, or for a complex system that may be supposed to have reached a steady state by collision processes. Thus the solutions may be right in just the conditions where they have been used, for just those quantities that remain constant in the motion. But this limitation of the conditions for the theory to give the right answers hardly justifies the attempt to give a special philosophy of quantum theory.

The difficulty is much less serious in classical mechanics, and even in general relativity, for in these it is possible to state hypotheses about the values of the variables for any t . The differential equations being supposed solved, the variables on each set of hypotheses at any other time could be calculated. The observed values of any functions of them could then be used to decide between the hypotheses by the principle of inverse probability. It would not even be necessary to make all the observations at the same time. But in quantum theory we must strictly allow for reactions between photons and matter, which will disturb the motions at intermediate times and prevent exact determinations at any one time, even given complete specification at any other. It would apparently be necessary to attach coordinates to every photon in the universe. I make no attempt to solve the problem, but I think it worth while to direct attention to what the problem is. A result quoted above from C. G. Darwin suggests that even in the Schrödinger system it is still possible to speak significantly of a coordinate and momentum having definite values simultaneously; for to speak of it is not the same as to say that we can measure them exactly. But it may happen that to state the hypotheses about the state of a system clearly we must reintroduce many ideas that present quantum theory rejects as "unobservable." However, the rejection of unobservables as usually stated covers a confusion similar to that in the second method of dimensions*. It has no logical force, but may lead to suggestions or to an economy of statement. It does not necessarily lead to the latter. Consider, for instance, what might have happened had Neptune had so low a reflecting power that it could not be seen. Astronomers would have had to decide between two

* 'Theory of Probability,' pp. 343-48; H. Dingle, 'Nature,' cxli. pp. 21-8 (1938)'

courses : to fit empirical terms to the extra perturbations of Uranus and Saturn, with no definite rule connecting them ; or to use the data to infer as nearly as possible the elements of the invisible planet and use them as a means of calculating future observations of the known ones. The first way would reject the unobservable, but would involve calculating every term irrespective of the others ; the second would make full use of information provided by each term about the others. The first, again, would require separate determinations of amplitude and phase for each term capable of being detected by observation ; the second would introduce as adjustable parameters the mass and orbital elements of the invisible planet, and nothing else. This is a case where the introduction of the unobservable would lead to a more useful and convenient expression of the probable positions of the known bodies, given their positions as observed to date. This will not always happen, but it is worth while to point out that it can happen, and that there is no practical or philosophical necessity for total rejection of, for instance, orbits within the atom even within the quantum theory. They would not be the classical orbits, but it remains possible that the quantum theory can be stated clearly only if it retains the idea of definite orbits.

XV. *The Propagation of Pressure Fluctuations of Large Amplitude in Air Columns.*

By L. J. KASTNER, M.A.*

[Received May 10, 1940.]

The Helmholtz-Kirchhoff Equation.

The velocity of propagation of pressure vibrations in air contained in tubes was investigated theoretically by Helmholtz † and Kirchhoff ‡, and the equation

$$u = a \left(1 - \frac{\gamma}{d\sqrt{\pi n}} \right) \quad (1)$$

was derived, in which a represents the velocity of sound in free air, d is the diameter of the tube, n the frequency of the vibration, and u the velocity of propagation in the air contained in the tube.

The constant γ , according to Kirchhoff, depends upon the friction and heat conductivity of the gas filling the tube, and can be calculated from the expression

$$\gamma = \sqrt{\mu} + \left(\frac{a}{b} - \frac{b}{a} \right) \sqrt{\nu}, \quad (2)$$

where $\mu = \frac{\eta}{\rho}$ and η = coefficient of viscosity of the gas,

ρ = density of the gas.

$\nu = \frac{K}{C\rho}$ a = true velocity of sound in free air,

b = "Newtonian" velocity of sound in free air.

$\frac{a}{b} = \sqrt{\frac{C'}{C}}$ K = coefficient of conductivity of the gas,

C' = specific heat of the gas at constant pressure,

C = specific heat of the gas at constant volume.

The equation (1) has been tested experimentally by several investigators§, and has been found to be approximately true, although the

* Communicated by the Author.

† H. Helmholtz, *Wiss. Abhandlungen*, i. p. 383.

‡ G. Kirchhoff, *Pogg. Annalen*. cxxxiv. p. 77 (1868).

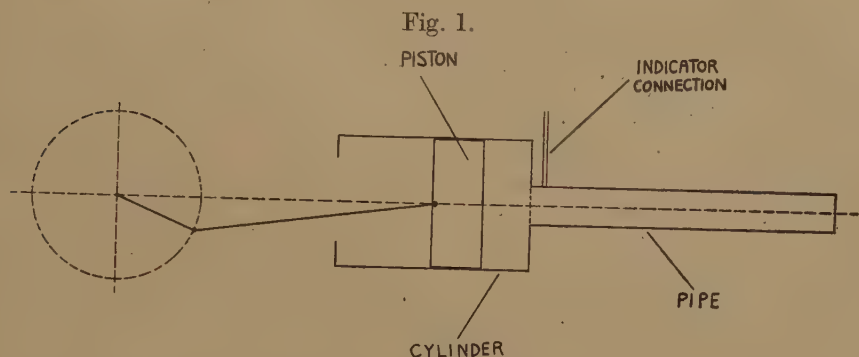
§ D. J. Blaikley, *Phil. Mag.* xvi. p. 447 (1883) and xviii. p. 328 (1884); J. Muller, *Annalen der Physik*. Bd. 11, p. 331 (1903); E. Gruneisen & E. Merkel, *Annalen der Physik*, lxvi. p. 318 (1921).

value of the constant γ has in some cases been found to vary considerably from that calculated from equation (2). The experimental investigations referred to, however, were carried out with vibrations of very small amplitude, and, so far as the writer is aware, little information exists concerning the behaviour of pressure fluctuations of large amplitude in air columns. A recent series of experiments on pressure waves in air provided an opportunity for examining the accuracy of the Helmholtz-Kirchhoff equation as applied to waves in which the average peak amplitude was about 5 inches of mercury above the mean pressure in the pipe.

At the outset it should be emphasized that no attempt is made here to provide a theory to explain the propagation of pressure vibrations of high amplitude in pipes. No more will be attempted than a comparison of the existing theory with the results of experiments.

Apparatus.

The initial pressure pulse from which the vibrations were built up was produced by a piston moving in a cylinder of 5 inches bore, the stroke



of the piston being 4.725 inches. The piston was driven through a connecting-rod and crankshaft assembly, the crankshaft being rotated by an electric motor. The cylinder was fitted with two valves in its head, the purpose of the first of these being to charge the cylinder with air, and of the second—hereafter to be called the exhaust valve—to eject the air through a pipe. The valve timing was so arranged that every fourth stroke of the piston the exhaust valve opened at the Bottom Dead Centre and remained open until the piston had reached its highest point, thus causing the charge of air in the cylinder to be expelled into a pipe, the far end of which was open to the atmosphere (see fig. 1).

As the piston moves upwards on the exhaust stroke it causes the air passing through the exhaust valve to be accelerated, and a pulse of pressure is produced which travels down the pipe, is reflected at the

open end and returns to its starting point, where it is again reflected. The process is continuous, and a wave is built up in the pipe from these reflexions.

The pressure wave so formed was measured by an indicator connected in the pipe close up to the exhaust valve—actually some 3 inches away from it. The indicator used consisted of a rotary valve so designed that the pressure in the pipe at the point where the indicator was connected could be measured directly on mercury columns at time intervals corresponding to 20 degrees on a crank-angle scale, the total duration of any pressure wave between successive openings of the exhaust valve being 4 strokes of the piston, or 720 degrees crank-angle. In this way, a pressure record could be obtained over the whole interval of time between successive openings of the exhaust valve.

The influence of pipe diameter was studied by using pipes of $1\frac{3}{8}$, $2\frac{3}{32}$, and $2\frac{7}{8}$ inches bore, the $1\frac{3}{8}$ and $2\frac{3}{32}$ inch pipes being constructed of solid drawn brass, and the $2\frac{7}{8}$ inch pipes of galvanized sheet iron *. For any given diameter the pipe length could be varied in six-inch increments from 0 to 144 inches. The speed of rotation of the crankshaft, and hence the speed of the piston, could also be varied by regulating the supply of power to the electric motor. For purposes of comparison, three standard speeds were chosen—1265 R.P.M., 1615 R.P.M., and 1992 R.P.M. In order to ensure accuracy in the measurement of speed, a solenoid-operated revolution counter was employed.

Experimental Results.

Specimen diagrams obtained by means of the indicator, and illustrating the nature of the pressure fluctuation in the pipe, are shown in fig. 2.

* It seems unlikely that this difference in pipe material will have much influence. If the velocity of the pressure fluctuations can be expressed approximately in the form

$$u = \frac{1}{\sqrt{\frac{\rho}{g} \left(\frac{1}{E} + \frac{1}{E_1} \frac{D}{t} \right)}},$$

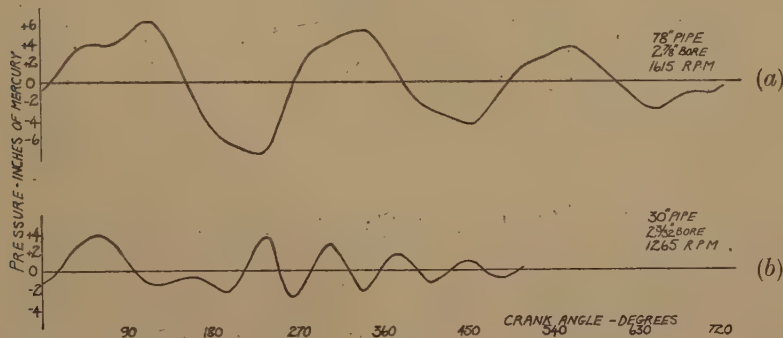
where E and E_1 are the moduli of elasticity of the fluid in the tube, and of the material of the tube respectively, and D and t refer to the mean diameter of the tube and the thickness of the wall, then if the fluid filling the tube be a gas, the

term $\frac{1}{E_1} \frac{D}{t}$ is unimportant compared with $\frac{1}{E}$ unless the thickness of the wall is very small compared with the diameter. For the pipes used the ratio $\frac{D}{t} < 100$ in all cases.

The diagrams given show clearly the differences between the type of vibration obtained with long- and with short-period pipes. With a short-period pipe the positive and negative reflexions of the initial pressure pulse follow one another in quick succession, and the vibration is degraded rapidly with time, the residual wave from the previous cycle being negligible. After the valve has closed and the piston is no longer in communication with the pipe, the free vibration will therefore be a smooth curve of regular shape. For a long-period pipe, however, the residual wave from the previous cycle will be strong, and may cause irregularities in the shape of the vibration, as shown in fig. (2 a).

After the exhaust valve has closed the vibration in the pipe is no longer affected by events in the cylinder, and settles down to a natural period corresponding to the length of the pipe, which is now to be considered as open at one end and closed at the other.

Fig. 2.



The natural frequency of such a vibration will therefore be $u/4L$, if u is the velocity of propagation of the wave in feet per second, and L is the pipe length in feet. If the total crank-angle occupied by one complete wave-length is known, the frequency can be expressed in terms of the speed of revolution of the crankshaft, that is, if N is the speed of revolution and p the total crank-angle occupied by one complete wave-length, the frequency is given by

$$n = \frac{6N}{p}$$

Therefore,

$$\frac{6N}{p} = \frac{u}{4L},$$

or if l is the pipe-length in inches

$$u = \frac{2Nl}{p}$$

Thus, if a series of diagrams be taken at known rotational speeds and with measured pipe-lengths, and if the natural period of the free vibration be observed from the indicated curves, the value of the velocity u can be obtained over a range of frequencies for any given pipe-diameter.

Tables I., II., and III. give the measured values of the natural period and the velocities of propagation derived from them for a number of pipe-lengths at the three standard rotational speeds for the three pipe-diameters investigated. The results are shown plotted in fig. 3.

It appears from fig. 3 that the velocity of propagation varies with the measured pipe-length in a different manner for each of the three pipe-diameters. This effect is attributable to the fact that the measured length of a pipe does not represent the true effective length*. There are two reasons to account for this. Firstly, it was necessary to provide an adaptor, roughly conical in shape and some six inches long, for connecting the pipes to the cylinder—the length of the adaptor being included in the measured length of any pipe—and the effective equivalent length of such an adaptor will not in general be equal to its measured length. Secondly, a correction will be required to account for the effect of the open end of the pipe.

Thus the true velocity of propagation for a pipe of measured length l inches will be given by

$$u = \frac{2N(l+x)}{p},$$

where x is the required additional length in inches, N and p having the same meanings as before. The following method was adopted for determining the value of x :

A number of values of the apparent velocity of propagation were taken from the mean curves of fig. 3, for measured lengths increasing by five-inch increments from 25 inches upwards. If u_a is the apparent velocity of propagation corresponding to a pipe of measured length l inches, then

$$u_a = \frac{2Nl}{p}.$$

Suppose an extra length x inches is now added to the measured length. We then have

$$u_x = \frac{2N(l+x)}{p}$$

or

$$\frac{u_x}{u_a} = \frac{l+x}{l}.$$

Hence a series of values of the velocity of propagation can be obtained, corresponding to any value of the additional length x . This process was

* A similar effect was noted by E. Gruneisen and E. Merkel (*loc. cit.*).

TABLE I.—1 $\frac{3}{8}$ " Bore.

Mean rotational speed = 1265 R. P. M.					Mean rotational speed = 1615 R. P. M.					Mean rotational speed = 1992 R. P. M.				
Measured pipe- length l ins.	Speed N R. P. M.	Mean natural period p° .	$u = \frac{2Nl}{p}$ ft./sec.	Measured pipe- length l ins.	Speed N R. P. M.	Mean natural period p° .	$u = \frac{2Nl}{p}$ ft./sec.	Measured pipe- length l ins.	Speed N R. P. M.	Mean natural period p° .	$u = \frac{2Nl}{p}$ ft./sec.	Measured pipe- length l ins.	Speed N R. P. M.	Mean natural period p° .
24	1267	78	780	30	1615	116	834	30	1991	148	807			
30	1274	93	822	36	1617	137	850	36	1993	168	854			
36	1268	108	845	42	1619	153	887	42	1990	192	871			
42	1274	118.5	903	54	1608	191	909	54	1997	243	886			
48	1266	133	914	60	1608	210	917	60	1997	261	918			
54	1267	150	912	68	1612	236	929	68	1996	290	936			
60	1265	165	919	72	1616	248	938	72	1987	304	941			
68	1270	184	938	84	1618	283	960	84						
78	1270	209	948	90	1616	300	969	90						
90	1262	234	971											
102	1259	263	976											
120	1271	304	1002											

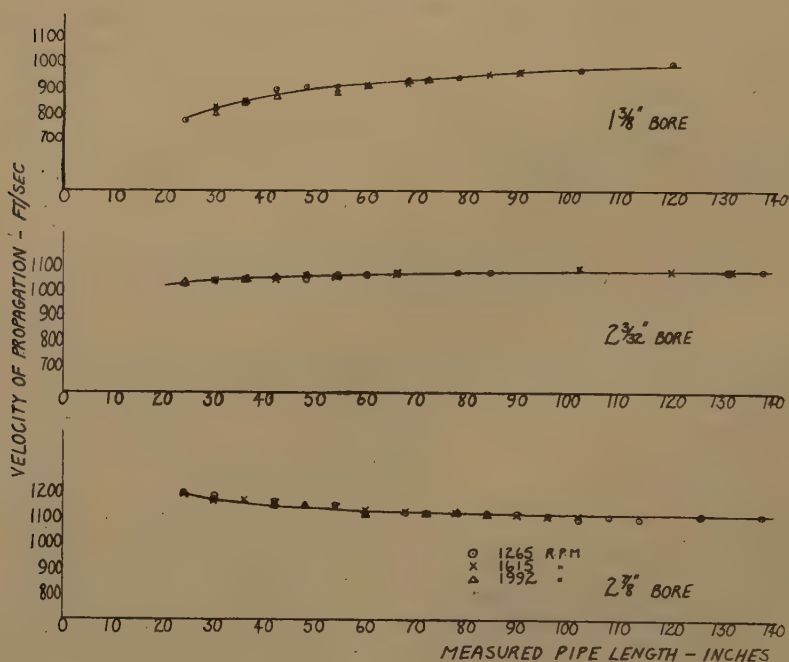
TABLE III. -- 27" Bore.

Mean rotational speed = 1265 R.P.M.				Mean rotational speed = 1615 R.P.M.				Mean rotational speed = 1992 R.P.M.			
Measured pipe- length l ins.	Speed N R.P.M.	Mean natural period p° .	$u = \frac{2Nl}{p}$ ft./sec.	Measured pipe- length l ins.	Speed N R.P.M.	Mean natural period p° .	$u = \frac{2Nl}{p}$ ft./sec.	Measured pipe- length l ins.	Speed N R.P.M.	Mean natural period p° .	$u = \frac{2Nl}{p}$ ft./sec.
30	1269	64	1190	24	1624	65	1200	24	1987	79	1206
42	1268	91	1170	30	1613	83	1168	30	1987	101	1180
54	1262	118	1155	36	1615	99	1174	42	1992	145	1152
60	1265	135	1124	42	1615	116	1170	48	1988	165	1156
68	1266	153	1126	54	1619	152	1150	60	1997	214	1120
72	1270	162	1129	60	1615	170	1140	72	1989	255	1121
78	1267	176	1123	68	1618	194	1132	78	1994	277	1122
84	1266	189	1124	72	1620	208	1121	84	1992	300	1116
90	1265	203	1124	78	1614	224	1124				
96	1269	219	1111	90	1616	262	1111				
102	1258	234	1096	96	1614	280	1107				
108	1263	246	1109	102	1611	296	1110				
114	1260	261	1100								
126	1265	286	1116								
138	1265	313	1115								

performed for the three pipe-diameters, for a number of values of the additional length. The apparent velocities for a zero additional length were taken from the appropriate mean curves of fig. 3. The results of the process are shown in figs. 4 and 5, plotted against the inverse of the square root of the frequency.

It will be seen that in all cases a series of curves is obtained, first converging towards and then diverging away from a given line. For one particular value of x the velocity of propagation appears to be

Fig. 3.



independent of the frequency over the range explored. The critical value of x giving this condition was obtained by plotting the difference between a number of pairs of points, corresponding to different frequencies, against the value of the additional length x . This results in a series of straight lines intersecting at a point. The point of intersection gives a value of x corresponding to a zero difference between any two points on the curve connecting u and $1/\sqrt{n}$.

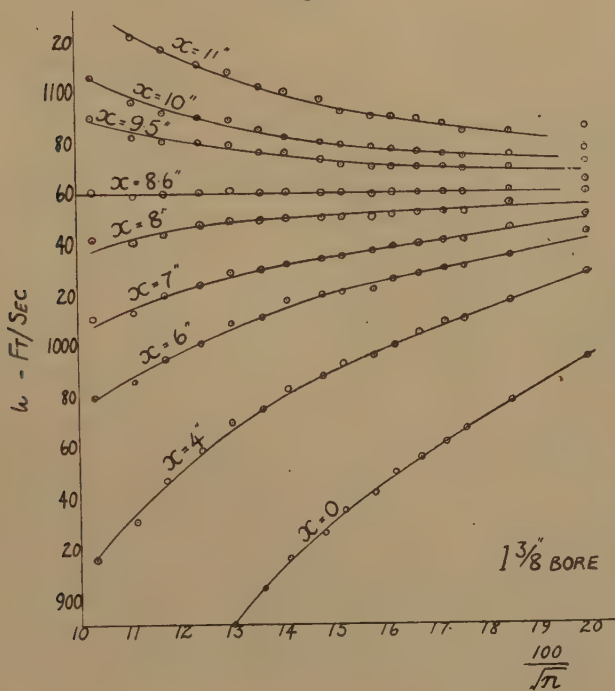
Tables IV., V., and VI. give for a number of pipe-lengths, at the three pipe-diameters investigated, the values of the velocity of propagation corresponding to the critical value of x , and also those values taken

from the mean curves of fig. 3 and corresponding to a zero additional length.

From the results given above, it is concluded that over the range of frequencies explored—from some 20 to 150 vibrations per second—the true velocity of propagation is independent of the frequency.

The true values of the velocity of propagation obtained from figs. 4 and 5 and the corresponding values of the extra equivalent length to be added

Fig. 4.



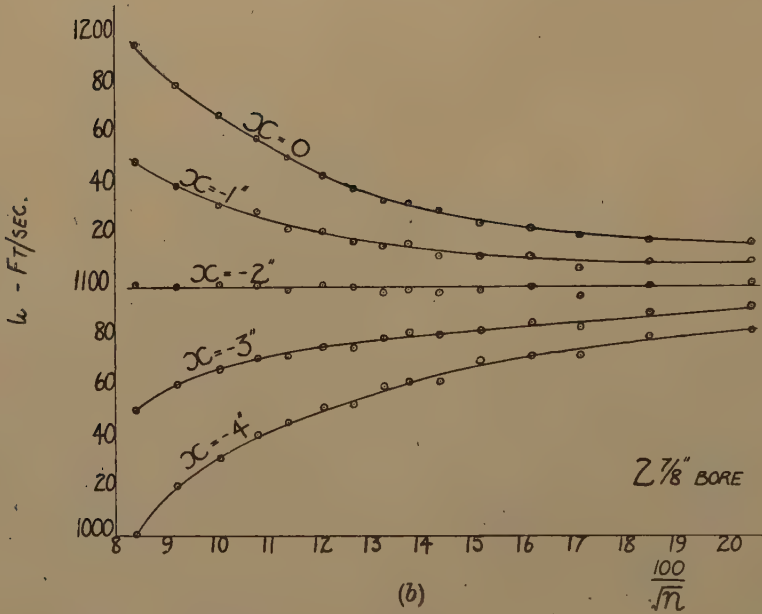
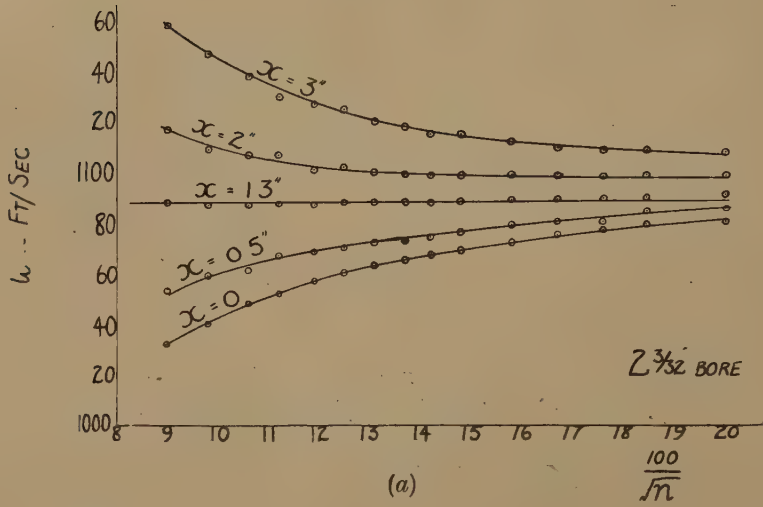
to the measured pipe-length are given below for the three pipe-diameters investigated.

According to equation (1), if the frequency n is constant

$$u_1 r_1 - u_2 r_2 = a(r_1 - r_2), \quad (3)$$

u_1, u_2 , being values of the velocity of propagation corresponding to pipes of radius r_1, r_2 . Since n appears to have no influence over the range explored, the values given in Table VII. may be substituted in equation (3) as a check.

Fig. 5.



The value of a , the velocity of sound in free air, at a temperature of t° Centigrade, is given by the formula

$$a^2 = a_0^2 \left(1 + \frac{t}{273} \right),$$

where a_0 is the velocity of sound in free air at 0° C. During the experi-

ments the temperature of the air in the pipe was approximately 24° C. If a_0 is taken as 1089 ft./sec.*, the appropriate value of a becomes

$$a = 1089 \sqrt{\frac{273+24}{273}},$$

$$= 1136 \text{ ft./sec.}$$

The results given above for the velocities of propagation in the $1\frac{3}{8}$, $2\frac{3}{32}$, and $2\frac{7}{8}$ inch pipes were grouped in pairs, and the value of a was

TABLE IV.— $1\frac{3}{8}$ " Bore.

Additional length (inches).	0	8.6	Frequency n vibrations per sec.	$\frac{100}{\sqrt{n}}$
Measured length (inches).	u_0 ft./sec.	$u_{8.6}$ (ft./sec.		
25	790	1061	94.8	10.3
30	822	1059	82.1	11.1
35	850	1060	72.9	11.7
40	872	1060.5	65.5	12.4
45	890	1061	59.5	13.0
50	904	1060	54.2	13.6
55	916	1060.5	50	14.1
60	926	1060	45.9	14.8
65	935	1060	43.1	15.2
70	942	1059.5	40.4	15.8
75	950	1060	38	16.2
80	956	1060	35.85	16.7
85	962	1060	34	17.2
90	967	1060	32.25	17.6
100	978	1061	29.35	18.5
120	995	1065	24.85	20.0

estimated from equation (3). The results are given in Table VIII., which should be compared with the value of 1136 ft./sec. given above.

To illustrate the way in which the velocity of propagation diminishes with pipe-diameter under the conditions investigated, the values of Table VII. were plotted against the reciprocal of the pipe-diameter (see fig. 6). For purposes of comparison, fig. 6 also includes values obtained by Blaikley (*loc. cit.*) for brass tubes of comparable diameter with vibrations of small amplitude, the frequency being 256 c.p.s. It appears

* 'Dictionary of Applied Physics,' R. Glazebrook, vol. iv. p. 686.

TABLE VI.—2 $\frac{7}{8}$ " Bore.

Additional length (x-inches).	0	-2	Frequency <i>n</i> vibrations per sec.	$\frac{100}{\sqrt{n}}$
Measured length (inches).	u_0 ft./sec.	u_{-2} ft./sec.		
25	1194	1099.5	143.2	8.36
30	1178	1098.5	117.6	9.22
35	1166	1099.5	99.8	10.02
40	1157	1099	86.6	10.8
45	1149	1097	76.5	11.4
50	1142	1099	68.55	12.1
55	1137	1098	62	12.7
60	1132	1096	56.6	13.3
65	1131	1097	52.2	13.8
70	1128	1096	48.4	14.4
80	1123	1096.5	42.1	15.4
90	1121	1098	37.4	16.4
100	1118	1094	33.5	17.3
120	1116	1098.5	27.9	19.0
140	1115	1099.5	23.9	20.5

TABLE VII.

Pipe-diameter (inches).	Velocity of propagation ft./sec.	Extra length (inches).
1 $\frac{3}{8}$	1060	+8.6
2 $\frac{3}{32}$	1089	+1.3
2 $\frac{7}{8}$	1098	-2.0

where

A_1 =amplitude of the original wave,
 A_n =amplitude of the n th reflexion,
 k_0 =reflexion coefficient for the open end,
 k_c =reflexion coefficient for the closed end,
 α =number of open end reflexions completed,
 β =number of closed end reflexions completed,
 L =distance travelled by the wave.

The attenuation factor m is given by the expression

$$m=\gamma\sqrt{\frac{n\pi}{ar}}, \dots\dots\dots (5)$$

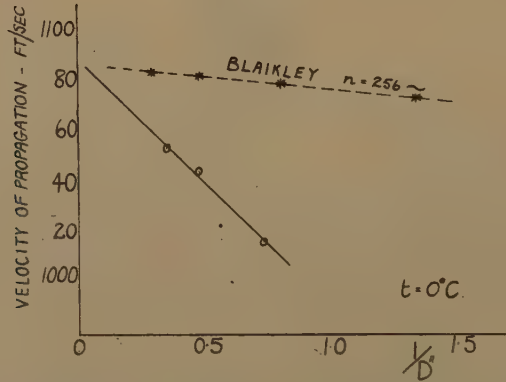
where γ , a , and n have the same meanings as in equation (1), and r is the radius of the pipe.

With large amplitudes the transference of heat will clearly be considerable, and the changes of density, in particular, which are consequent

TABLE VIII.

Pipe diameters (inches).	u_1 ft./sec.	u_2 ft./sec.	a ft./sec.
$1\frac{3}{8}, 2\frac{3}{32}$	1089	1060	1144.4
$1\frac{3}{8}, 2\frac{7}{8}$	1098	1060	1132.7
$2\frac{3}{32}, 2\frac{7}{8}$	1098	1089	1122.2
Average value of $a=1133.1$ ft./sec.			

Fig. 6.



on the pressure changes, will make the calculation of the constant γ —according to equation (2)—highly problematical. Further, the results of the present series of experiments have shown that under the given conditions the velocity of propagation is independent of the frequency. Hence equation (1) is not applicable and the validity of equation (5) is open to question, since equations (1) and (5) are to some extent interdependent. The degree of damping produced by attenuation and by reflexion from the ends of the pipe must therefore be found experimentally.

If the amplitude of the fundamental pulse of pressure produced by the piston at a crank-angle θ equals $f(\theta)$, and if one complete transit from the closed end to the open end of the pipe and back again occupies

$2L/u$ seconds (L here representing the length of the pipe)

$=12 NL/u$ degrees

$=p/2$ degrees, where p is the natural period of the vibration, the pressure at the closed end is given by the expression

$$P=f(\theta)-k_0(1+k_c)e^{-2mL}f(\theta-p/2)+k_0^2k_c(1+k_c)e^{-4mL}f(\theta-p) \\ -k_0^3k_c^2(1+k_c)e^{-6mL}f(\theta-3p/2)\dots\text{etc.}, \quad (6)$$

the second term of this expression referring to the first double reflexion of the fundamental pulse, and subsequent terms referring to successive reflexions of this. Further, if the net pressure at any instant during the free vibration (*i. e.* after the valve connecting cylinder and pipe has closed) is made up of the algebraic sum of the amplitudes of the

q th, $(q+1)$ th, $(q+2)$ th, $\dots$ $(q+r)$ th

reflexions, it can be shown that a point on the free vibration falling one whole wave-length later will be affected by the

$(q+2)$ th, $(q+3)$ th, $(q+4)$ th, $\dots$ $(q+r+2)$ th

reflexions only, and hence the amplitude ratio between the two points considered will be given by the relation

$$A_2/A_1=\{k_0k_ce^{-2mL}\}^2,$$

where A_1 , A_2 refer to the amplitudes of the net pressure wave at the points considered. Thus, a measurement of the amplitudes one wave-length apart affords a method of finding the reduction factor for one complete transit.

This is still true if a residual wave from the previous cycle is present. In this case two waves must be considered, the natural period " p " and the factor $k_0k_ce^{-2mL}$ being the same for both, but there may be a phase difference between them.

Then if A_2/A_1 be the ratio between successive points one wave-length apart for one wave, and if a_2/a_1 be the corresponding ratio for the other wave at the same points,

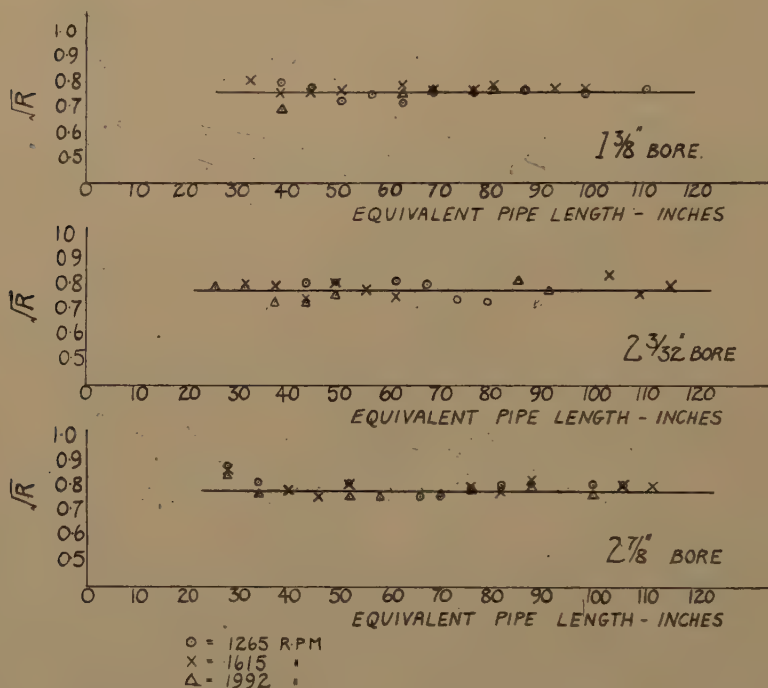
$$A_2/A_1=a_2/a_1=\{k_0k_ce^{-2mL}\}^2.$$

Therefore

$$\frac{A_2+a_2}{A_1+a_1}=\frac{a_2}{a_1}=\{k_0k_ce^{-2mL}\}^2, \text{ as before.}$$

The square root of the amplitude ratio between successive peaks or successive troughs, which should correspond to the factor $k_0 k_c e^{-2mL}$, is shown plotted in fig. 7 for the three pipe-diameters investigated. From the figures it will be seen that the factor is roughly constant for a given pipe-diameter, and shows no sign of falling off as the pipe-length is increased. The irregularities corresponding to small pipe-lengths may be due to the fact that for such pipes the angular interval between successive indicator readings—20 degrees—crank—angle may be a fairly considerable fraction of the natural period of vibration.

Fig. 7.



The points plotted in fig. 7 were obtained from a number of indicated pressure waves, similar to those of fig. 2. The amplitude values used refer only to the free vibration of the pressure waves, and no values are included corresponding to points occurring before the instant of closing of the valve connecting cylinder and pipe, so that the influence of the fundamental pulse is eliminated. With long-period pipes a single free vibration will occupy a considerable proportion of the total time between successive openings of the valve, so that values are not given for pipes longer than about 120 inches, owing to the uncertainty of the experimental results.

TABLE IX.—1 $\frac{3}{8}$ " Bore Pipes.

Measured pipe- length (inches).	Effective pipe- length (inches).	A ₀ ins. Hg.	A ₁ ins. Hg.	A ₂ ins. Hg.	A ₃ ins. Hg.	$\frac{A_2}{A_0}$	$\frac{A_3}{A_1}$	$\sqrt{R}$.
Mean rotational speed=1265 R.P.M.								
30	38.6	5.5	4.3	3.4	2.8	0.62	0.651	0.798
36	44.6	6.6	4.9	4.0	3.0	0.606	0.614	0.78
42	50.6	8.4	6.2	4.2	3.4	0.5	0.55	0.725
48	56.6	8.6	6.4	4.8	..	0.56	..	0.748
54	62.6	9.6	7.9	4.9	..	0.51	..	0.714
60	68.6	10.6	8.8	6.0	..	0.566	..	0.752
68	76.6	10.7	9.0	5.9	5.3	0.55	0.59	0.755
78	86.6	11.6	9.4	6.5	5.7	0.56	0.607	0.764
90	98.6	10.6	9.6	6.0	5.3	0.566	0.553	0.748
102	110.6	9.8	8.8	5.8	..	0.591	..	0.77
Mean rotational speed=1615 R.P.M.								
24	32.6	6.2	4.85	4.0	..	0.645	..	0.803
30	38.6	8.4	6.6	4.5	3.9	0.536	0.59	0.75
36	44.6	8.6	7.0	4.9	..	0.57	..	0.755
42	50.6	11.4	9.2	6.6	..	0.58	..	0.762
54	62.6	11.0	10.0	6.8	6.1	0.619	0.61	0.784
60	68.6	11.8	10.4	6.8	6.2	0.576	0.596	0.765
68	76.6	..	9.9	6.5	5.8	..	0.586	0.765
72	80.6	9.6	10.0	6.6	5.4	0.688	0.54	0.784
84	92.6	9.8	9.7	5.8	..	0.592	..	0.77
90	98.6	10.7	9.7	6.4	..	0.598	..	0.773
Mean rotational speed=1992 R.P.M.								
30	38.6	11.6	10.2	5.5	..	0.474	..	0.689
54	62.6	12.3	11.5	7.4	6.2	0.601	0.54	0.755
60	68.6	11.8	11.0	7.0	..	0.592	..	0.77
68	76.6	12.1	11.0	7.0	..	0.579	..	0.76
72	80.6	12.6	11.6	7.5	..	0.595	..	0.771
78	86.6	12.6	11.4	7.3	..	0.58	..	0.762

Table IX., which refers to pipes of 1 $\frac{3}{8}$ inch bore, is included to illustrate the magnitude of the pressure fluctuations. The columns headed A₀, A₁, A₂, A₃ give the amplitudes, in inches, of mercury above or below

atmospheric pressure of successive peaks and troughs on the indicated pressure waves. Hence the number A_2/A_0 , or A_3/A_1 , will be the amplitude ratio between points separated by one complete wave-length. If the mean ratio for any wave be R the quantity $\sqrt{R}$ should correspond to the factor $k_0 k_e e^{-2mL}$ for the pipe considered.

It is advantageous to obtain the ratio for points separated by a whole wave-length, rather than by half a wave-length, since this reduces the possibility of any error which might arise if the mean pressure in the pipe were at any time greater than atmospheric.

The constancy of the reduction factor appears to show, for pipes of moderate length, that the attenuation factor m is unimportant compared with the product of the reflexion coefficients, even when the amplitudes of the vibrations are large. Throughout the series of experiments the greatest pipe-lengths used were kept well below that value which would resonate with the natural frequency of the piston, and no unusual effects were noticed which could be attributed to harmonics of that frequency.

A direct determination of the constant m would necessitate a provision for measuring the pressure in the pipe at a number of points along the length of the pipe itself, and the correct phasing of the indicator for each of these points and the elimination of residual wave effects would be a difficult matter. For this reason no attempt was made to measure the attenuation constant directly, since it was felt that experiments on the lines indicated would be unlikely to give satisfactory results with the apparatus available.

Conclusions.

The results of the investigation indicate, within the scope of the experiments described, that the velocity of propagation of pressure waves in air produced in a pipe excited by a piston is independent of the frequency, when the peak amplitude of the free vibration corresponds to a pressure of a few inches of mercury, and the number of vibrations varies between 20 and 150 per second. Under the above conditions the velocity of propagation appears to diminish at a greater rate with decrease of pipe-diameter than for cases when the amplitude is very small.

The waves generated in the pipe are attenuated as they travel along the pipe, and are also reduced in amplitude on reflexion from the ends of the pipe. For pipe-lengths up to 10 feet the attenuation constant appears to be unimportant compared with the open-end and closed-end reflexion coefficients, and the product of these latter coefficients does not vary greatly with changes of pipe-diameter.

XVI. *Magnetization in Crystalline Media.*

By Prof. W. PEDDIE *.

[Received November 18, 1940.]

A VERY large amount of experimental information has now accumulated regarding the phenomena of magnetization in crystals; and there is on the whole, good agreement amongst the results of different observers. There can thus be very little doubt that observed differences arise mainly from the unavoidable imperfection of crystals. But the number of individual crystals which have been experimented upon is not so great as to make it fairly certain that we can predicate, from the mean of observed results under given conditions, the true law of action in an ideal crystal. Yet it is necessary to know what that ideal action should be; and that can only be determined from theory.

Various theories are possible, and have been utilized. Apart from the correspondence of results with observation, we can only discriminate between these theories on the ground of simplicity of fundamental postulates. The simplest of all is the theory originated, almost a century ago, by W. Weber. It postulates that the magnetic elements in atoms constituting a magnetic body respond freely to the directional influence of an outside field, except in so far as the mutual interactions of these elements give rise to an internal field.

This statement of course does not preclude the influence of other forces in altering the relative positions of atoms, and so affecting the internal field or disturbing its action. The effect of these other forces are theoretically calculable when the necessary data regarding them are given, whether they be electrical or mechanical in their origin.

The influence of crystalline boundaries also is a most important case in point. When the data are given regarding these, the results are also theoretically calculable; and the experimenter aims at the practical reduction of their influence. So also does the theorist, in the first place, by carrying the boundaries to infinity.

Maxwell first dealt with the question of instability in magnetic orientation. For its simplification in the matter of expression, he introduced a non-magnetic action. Others followed similarly. Ewing dealt with this

* Communicated by the Author.

point specially, using magnetic interaction alone. Subsequent experimental development has been large; and the results in general sustain Weber's theory in a striking manner. That this should be so is not entirely self-evident; for the existence of electrical dipoles in the atoms could give rise to an internal electrical field similar to Weber's magnetic internal field. This action may indeed be explanatory of some peculiarities in the magnetic behaviour of certain crystals.

Doubts have been expressed as to the adequacy of Weber's theory, especially in regard to hysteresis, and so suggest the necessity of investigation.

The following discussion amounts only to a brief summary of main results, which I hope may be found of use.

The Internal Field in Cubic Crystals.

For the purpose of carrying out a discriminative comparison between theory and observation, it will be sufficient to limit the discussion to the case of crystals belonging to the cubic system to which the majority of magnetic crystals belong; and, further, in these, to the case of magnetization parallel to a face plane of a homogeneous crystal unlimited in size.

Crystalline magnetic æolotropy is made evident by the fact that the direction of magnetization does not in general coincide with the direction of the external magnetic field which give rise to it. This means that there is a component of an internal field, arising from the magnetization, which is transverse to the direction of the magnetization. This transverse component is known to be expressed, in terms of the angle which the direction of magnetization makes with that of the nearest crystalline axis of stable magnetization, by the formula (Proc. Roy. Soc. Edin. 1905, p. 1039)

$$T = A \sin 4\theta,$$

where A is a constant which is numerically equal to the maximum value of the transverse component of the internal field.

The condition of equilibrium under an external field H , directed at an angle θ_0 to the stable axis, is

$$H \sin (\theta_0 + \theta) = A \sin 4\theta.$$

If we express the magnitude of the field H in terms of A as a unit through the condition $H = Ah$, the equation of equilibrium becomes

$$h \sin (\theta_0 + \theta) = \sin 4\theta. \quad \dots \dots \dots (1)$$

Under the condition $\partial h / \partial \theta = 0$ the equilibrium becomes unstable and the magnets seek a new position of equilibrium in nearer alignment to the field h . The state of instability therefore requires

$$h \cos (\theta_0 + \theta) = 4 \cos 4\theta. \quad \dots \dots \dots (2)$$

From these two equations we obtain

$$\sin 4\bar{\theta} = \sqrt{\frac{16 - \bar{h}^2}{15}}, \quad \cos 4\bar{\theta} = \pm \sqrt{\frac{\bar{h}^2 - 1}{15}}, \quad \dots \quad (3)$$

where $\bar{\theta}$ is the breaking angle, the angle of instability, at which $h = \bar{h}$.

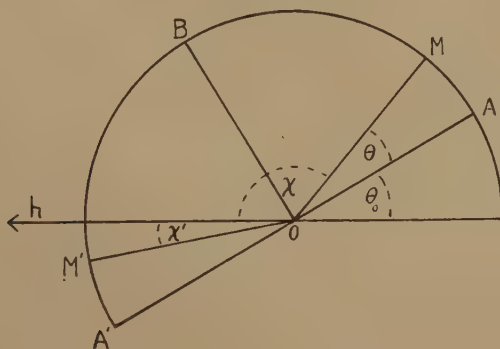
The Curves of Magnetization.

From equations (3) it appears that no break can occur except in the circumstances

$$h < 1 > 4.$$

If h be less than unity, the force is unable to pull the magnets out to the breaking angle. If h be made greater than 4 all the magnets have

Fig. 1.



already passed through the unstable stage and have been re-aligned in the neighbourhood of the direction of h .

Consider now the direction OA in fig. 1 to be that of a stable axis of the crystal, making an angle θ_0 with the fixed direction reverse to h . When h acts, magnets aligned along OA are deviated from it through an angle θ and take up the direction OM , so that (1) becomes

$$h \sin (\pi - \theta_0 - \theta) = \sin 4\theta, \quad \dots \quad (4)$$

$$\text{or} \quad h \sin \chi = -\sin 4(\theta_0 + \chi), \quad \dots \quad (5)$$

where χ is the angle between the direction of h and that of magnetization.

Now if we consider a volume of the magnetically saturated crystal which possesses unit magnetic moment, we get $\cos \chi = \sigma$ as the value of the magnetic moment induced in the direction of h . Thus

$$h \sqrt{1 - \sigma^2} = -\sin 4(\theta_0 + \cos^{-1} \sigma). \quad \dots \quad (6)$$

This is the equation of the (σ, h) curve of magnetization of an infinite

homogeneous cubic crystal magnetized in a plane parallel to a face plane of the cubic lattice.

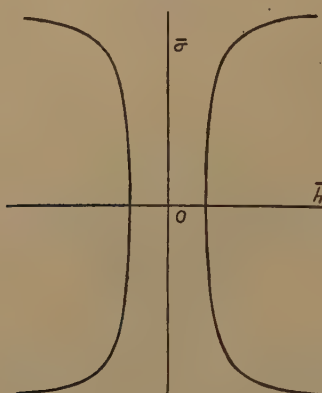
With a given value of θ_0 , (6) gives the value of h corresponding to any chosen value of σ (which must lie between the limits ± 1) until the instability point is reached. At that stage we have

$$\bar{h} \sin \bar{\chi} = \sin 4\bar{\theta},$$

and so, by (3),

$$\begin{aligned} \bar{\theta}_0 &= \pi - \bar{\theta} - \sin^{-1} \frac{\sin 4\bar{\theta}}{\bar{h}}, \\ &= \pi - \frac{1}{4} \sin^{-1} \sqrt{\frac{16 - \bar{h}^2}{15}} - \sin^{-1} \sqrt{\frac{16 - \bar{h}^2}{15\bar{h}^2}} \dots \dots \dots (7) \end{aligned}$$

Fig. 2.



This gives the constant value of θ_0 corresponding to any chosen value of $\bar{h}$ between the limits 1 and 4.

The value of $\bar{\sigma}$ is

$$\bar{\sigma} = \pm \sqrt{\frac{\bar{h}^2 - 1}{15\bar{h}^2}}, \dots \dots \dots (8)$$

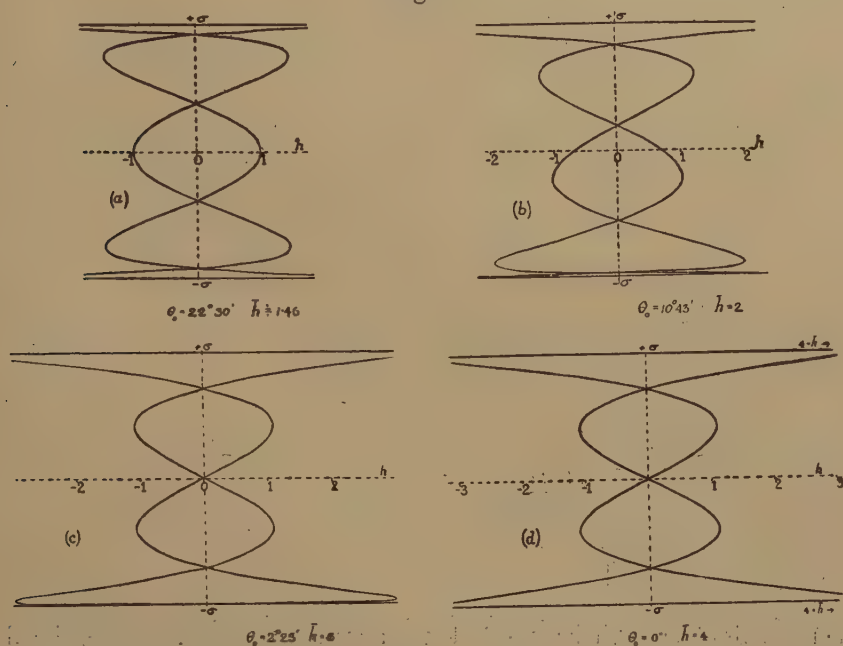
the positive or negative sign being used according to χ exceeds or falls short of $\pi/2$. This is the locus, fig. 2, of instability points or turning points on the magnetization curve. In particular, at $\bar{h} = \pm 4$, $\bar{\sigma} = \pm 1$.

Equation (6) enables us to plot the magnetization curve from $h=0$ to $h=4$, and beyond that to $h=\infty$.

The form of (6) shows that the full magnetization curve is symmetrical about the axis of h and also about the axis of σ . Four of these curves are shown in fig. 3, where, for the sake of clearness, the reflexion in the line $\sigma=0$ is not drawn for the curves (b) and (c). The curves (a) and (d)

coincide with their own reflexions. In (b) and (c) the maximum turning point is drawn on the side of negative σ .

Fig. 3.



The following Table, giving values of $\bar{h}$ in the first column, gives, in the last four columns, values of σ at crossing points where $h=0$. These therefore represent values of residual magnetization:—

1.0000	.9239	.3827	—3827	—9239
1.0010	.9374	.4166	—3489	—9091
1.0060	.9582	.4752	—2861	—8798
1.0100	.9673	.5046	—2537	—8634
1.0200	.9799	.5519	—1994	—8339
1.0400	.9962	.6426	—0875	—7662
1.0800	.9998	.6936	—0189	—7204
1.0886	1.0000	.7071	—0000	—7071
1.1000	.9997	.6911	—0224	—7228
1.2000	.9948	.6314	—1019	—7754
1.3000	.9627	.4895	—2705	—8721
1.4000	.9329	.4048	—3603	—9144
1.4500	.9273	.3910	—3743	—9239
1.4652	.9239	.3827	—3827	—9239
1.5000	.9313	.4009	—3643	—9162
1.7000	.9606	.4828	—2779	—8758
2.0000	.9826	.5634	—1858	—8262
2.5000	.9956	.6376	—0940	—7703
3.0000	.9991	.6897	—0243	—7241
3.5000	.9999	.6961	—0125	—7159
4.0000	1.0000	.7071	—0000	—7071

If we consider passage from a higher to a lower stable state, we see that the second largest crossing-point (last column in the Table on p. 229) may be omitted. For, although its data correspond to a stable magnetization, it does not give the most stable condition attainable by the magnets when they are pulled, with reversal of the sign of the magnetization, through a position of instability. The energy gained in reaching that position of negative magnetization from the higher position of positive magnetization is greater than the energy needed to extract the magnets from it. Thus the magnets tend to settle, when the field is removed, at the negative σ numerically equal to the positive σ given in the second column of the Table. So we only need to consider a diagram exhibiting positive σ in the region containing the largest $\bar{h}$ along with its reflexion in the axis of h . All the data in the middle range of the Table, between $\bar{h}=1.0886$ and $\bar{h}=1.4652$, may be omitted. The region of small hysteresis corresponds to the earlier part of the Table and the region of large hysteresis corresponds to the later part. The intermediate fields only bring about passage from the smaller to the larger hysteresis areas. They correspond to values of ${}_1\bar{\theta}_0$ lying between $\pi/8$ and $\pi/4$, and of ${}_2\bar{\theta}_0$ from $\pi/2$ to $5\pi/8$. [There are *two* values of $\bar{\theta}_0$ corresponding to any given $\bar{h}$. These are equally inclined to the direction $3\pi/8$ measured from the direction opposite to that of h .]

In the general case the largest and the smallest values of $\bar{h}$ appear at turning points in both halves of the diagram with reference to the line $\sigma=0$.

Hysteresis.

So long as h is less than unity, no hysteresis occurs. At $h=1$, $\bar{\theta}_0=3\pi/8$, the loss of energy through hysteresis becomes suddenly finite. Its amount can be exhibited diagrammatically by dropping (fig. 3 (b) inverted) a perpendicular from the smaller turning point corresponding to the given value of $\bar{\theta}_0$, say at $h=-\bar{h}_1$, to the σ axis; doing the same at $h=+\bar{h}_1$, and prolonging the latter perpendicular upwards till it reaches, say at $\sigma=\sigma'_1$, the positively sloped branch of the magnetization curve proceeding from the point $h=-\bar{h}$. Twice the area included by this branch, the two tangents at $h=\pm\bar{h}_1$, and the axis of h ($\sigma=0$) gives, in the reduced (h, σ) units, the amount of the loss.

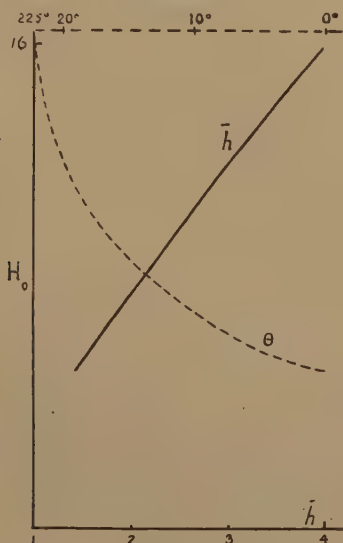
If we now, beyond $\sigma=\sigma'_1$, increase $\bar{h}$ from $\bar{h}_1$ to any point short of $h=\bar{h}_2$, where $\bar{h}_2$ is the largest value of $\bar{h}$ at the given $\bar{\theta}_0$, and then reduce it, the path is retraced to $h=+\bar{h}_1$. As h falls further to zero, the curved path is followed to the residual σ (as exemplified in the third column of the Table above). The curved stable path is still followed as h becomes

increasingly negative until $h = -\bar{h}_1$. Instability arises there, and the positive magnetization suddenly becomes negative and equal to $-\bar{\sigma}_1$. Half of the smaller hysteresis loop is thus traced.

A similar process holds in the case of any other $\bar{h}$ corresponding to a different $\bar{\theta}_0$.

If, by further increase of h from 0 to beyond $\bar{h}_1$, the point $h = \bar{h}_2$ is stably reached, instability again sets in, and σ suddenly takes the value σ_2' , where the tangent at the turning point meets the still higher positively sloped stretch of the magnetization curve. Further increase of h towards

Fig. 4.



infinity makes σ approximate to saturation ($\sigma=1$). On reduction of h to zero (highest crossing-point of the Table), and finally reversal of it, instability enters at $h = -\bar{h}_2$, and the magnetization then reverses to a value $-\sigma_2'$, and so on. The larger hysteresis loop is thus traced, the smaller one ceasing to be traced in the cyclical variation of h . The largest possible area has the value 16 in (h, σ) units. It is the rectangle $(\pm 4, \pm 1)$.

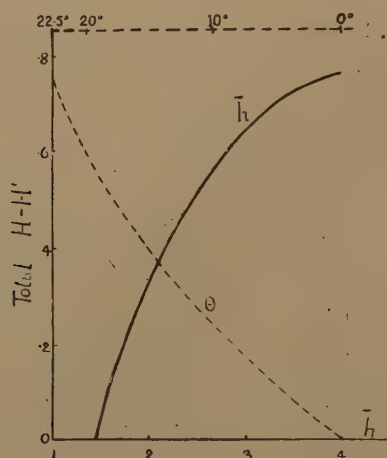
To find the total loss of energy, by hysteresis, manifested as a congeries of cubic crystals, having two of the cubic axes uniformly distributed in one fixed plane containing the direction of the external field, we can proceed as follows. Take the product of the amount of the hysteresis loss per molecule, having a stable axis in any direction θ_0 , into the number

of molecules distributed in the small range $d\theta_0$; and, in a new diagram having θ_0 as abscissa, erect an ordinate, whose length is numerically equal to that product, at the mid-point of the range $d\theta_0$. The area included by the locus thus found, the line of abscissæ, and any two ordinates, give the total loss of energy in the corresponding range of θ_0 , and therefore in that of $\bar{h}$.

The ascending and descending curves in fig. 4 show respectively the values of the hysteresis loss of energy, H_0 , per unit saturation magnetic moment, at corresponding values of $\bar{h}_1$ and ${}_1\bar{\theta}_0$.

The ascending and descending curves in fig. 5 show, similarly, the values of the excess of the total loss of energy, H , over the amount, H' , exhibited at $\bar{h}=1.4652$, ${}_1\bar{\theta}_0=\pi/8$.

Fig. 5.



The Curve of Instability.

Equation (7) is the polar equation of a curve representing the relation between h and θ_0 at the points of instability, the radius vector of the curve being made equal to $\bar{h}$, while θ_0 is, as before, measured from the reverse direction of h . One branch of the curve is drawn in fig. 6. It touches the circle of unit radius at the angular position $3\pi/8$; touches the radii $\theta_0=0$ and $\theta_0=3\pi/4$ at $h=4$; and is symmetrical about the shortest radius, $h=1$. The common value of h at $\theta_0=\pi/4$ and $\theta_0=\pi/2$, is 1.0887 in the reduced units—reduced so as to give results applicable to any crystal, whatever be the value of the maximum transverse field.

Fig. 6.

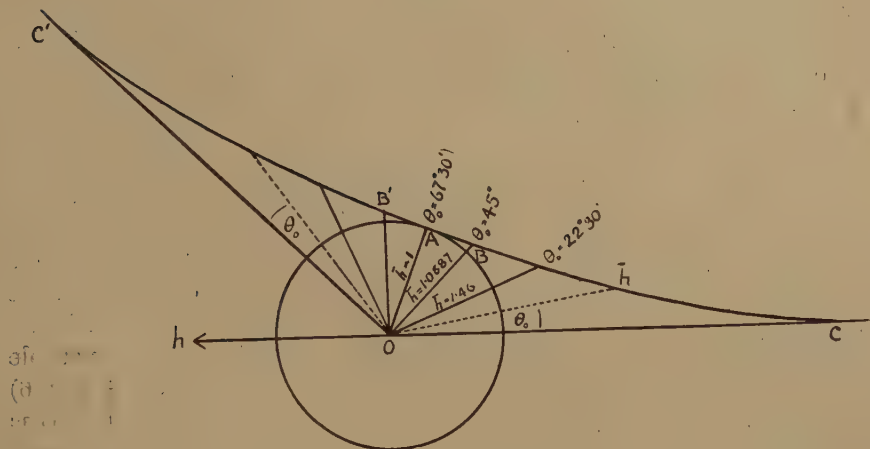
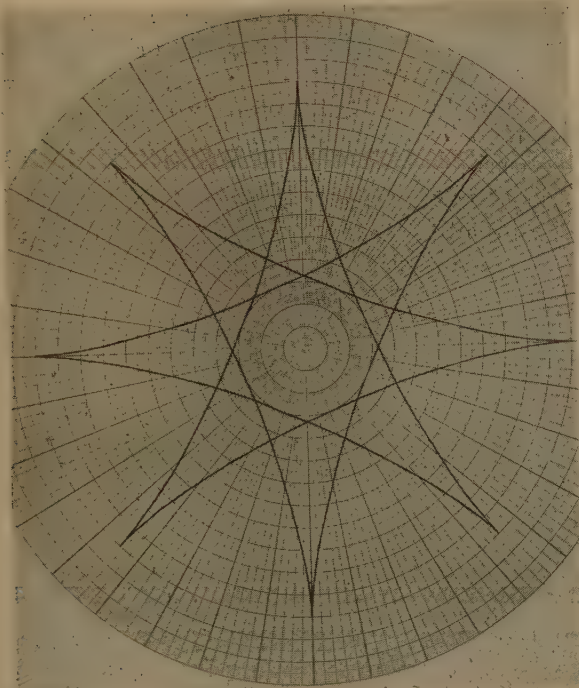


Fig. 7.



By the symmetry of the crystal it is evident that the curve repeats itself in each successive total angle of $3\pi/4$. So it becomes re-entrant after three complete revolutions of the radius vector (fig. 7). There are thus three values of h which correspond to any one direction of the stable axis, but belong to three separate orientations of the field, relatively to that axis, differing by $\pi/4$ successively.

Residual Magnetization.

The curve of instability shows that no sudden break in the orientation of a magnet at a given value of h can take place so long as that value lies within the limits $h=0$, $h=1$. If the field be removed, the magnet returns to its initial orientation, and there is no residual magnetization.

But, if h exceeds unity, while not exceeding 1.0887, the unstable position of the magnet passes outwards both ways equally from A (fig. 6) until the range of directional instability becomes $BOB'=\pi/4$. If OB is the original direction of the crystalline axis, OC', the direction of the axis perpendicular to OB is the position to which the magnet sets after break when h is removed. If OB' is the original direction, the final one is Oh. Always, a magnet, breaking from an axis in the region BOB', resets in relation to the perpendicular axis in the region C'Oh.

In fact, when $h=1.0887$, all the magnets originally distributed uniformly in the second positive octant, counting from the reverse direction to h , are, after removal of h , reset uniformly in the fourth octant, where their magnetic effect balances that of the magnets in the first octant. Similarly in the case of the magnets in the negative octants. Thus the total residual effect is that due to a uniform distribution through the two quadrants adjacent to the direction of h . So the residual magnetization is given by the expression

$$2m \int_0^{\pi/2} n \cos \theta d\theta,$$

where m is the moment of each individual magnet, and n is the density of distribution of the magnets per unit angle. If n_0 be the total number of magnets corresponding to unit saturation moment, we have $2\pi n=n_0$. The residual magnetic moment is therefore mn_0/π . That is to say, it already amounts, at $h=1.0887$, to the fraction $1/\pi$ of that obtainable by an infinite field.

Similar considerations with regard to the curve of instability points show that, as h increases from 1.0887 to 4, the breaking points move out along BC and BC' symmetrically till, at $h=4$, the end points are reached. All the magnets have then passed through the unstable stage and are aligned within the two octants adjacent to h , and are uniformly oriented throughout them when h is removed. Since all the magnets are

now crowded into one quarter of their original range, the residual moment has the magnitude

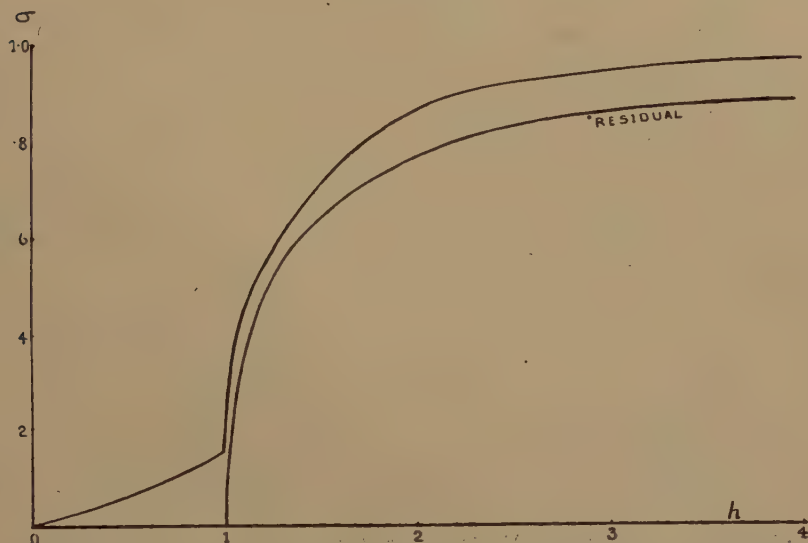
$$\frac{2\sqrt{2}}{\pi} mn_0,$$

which is almost exactly nine-tenths of its maximum value.

The treatment proceeds on the same lines when the field is not limited to the special values 1.0887 and 4. In all cases the resultant magnetization is in the direction of h by symmetry.

In initial magnetization from the random distribution of magnets, as h increases from 1.0887 to 1.4652 the numerical value of θ_0 (fig. 6)

Fig. 8.



proceeds from $\pi/4$ to $\pi/8$, and the reoriented magnets arrange themselves along the stable axes in the first negative octant adjacent to the field direction, spreading inwards from its outer boundaries towards its centre. Similarly, in the case of the magnets originally oriented in the part of the diagram below the line OC, the new positions spread in from the outer boundaries of the octant C'Oh. Thus, when the field is removed after $h=1.4652$, the reoriented magnets are distributed uniformly in the quadrant bisected by Oh, and the residual magnetization becomes $mn_0 \cdot 2/\pi$, in so far as reset groups act.

As h further increases from 1.4652 to 4, θ_0 changes from $\pi/8$ to zero, and the new stable axes for the reoriented magnets spread outwards from the centre of the first negative octant until, when h reaches the value 4 and is then reduced to zero, they are spread uniformly throughout that octant.

These conditions enable us to obtain readily the general expression for the residual moment corresponding to any value of h in the interior of any one of the three stretches just considered. The graph is shown in fig. 8.

Magnetization of a Randomly Crystallized Medium.

In a randomly crystalline medium, such as soft iron demagnetized by heating, the individual crystals may be regarded as having their axes directed in space with a randomly uniform distribution in all directions. Because of the complexity of the relation between magnetization and magnetizing force the general investigation of the problem is practically excluded. But it may be presumed that the magnetic action will in general present features similar in type to those which would be exhibited if we could build up a medium in which the crystals, otherwise randomly arranged, had one of the three sets of face planes in a cubic lattice always perpendicular to one fixed direction in space. The theory of magnetization in this case is readily treatable.

If we measure both θ_0 and χ from the direction of h , these angles giving respectively the direction of the stable axis and that of the magnetization, the equilibrium condition is

$$h \sin \chi = \sin 4(\theta_0 - \chi) = \sin \psi.$$

And the magnetic moment is

$$\frac{mn_0}{\pi} \int_0^\pi \cos \chi \, d\theta_0.$$

On integration this becomes

$$\frac{mn_0}{\pi} \left[\sin \chi + \frac{1}{4h} E(\chi, \alpha) - \frac{1-h^2}{4h} F(\chi, \alpha) \right]_{\theta_0=0}^{\theta_0=\pi},$$

where $\sin \alpha = h$, with h not greater than unity, and E and F are the Jacobian elliptic integrals. When h exceeds unity, the expression takes the form

$$\frac{mn_0}{\pi} [\sin \chi + \frac{1}{4} E(\psi, \beta)],$$

where $\sin \beta = h$, $\psi = 4(\theta_0 - \chi)$, and the limits of integration are those for the stretches $\theta_0 = 0$ to $\theta_0 = {}_1\bar{\theta}_0$ and $\theta_0 = {}_2\bar{\theta}_0$ to $\theta_0 = \pi$. As already stated, the two values of $\bar{\theta}$ are symmetrical about $3\pi/8$, so that the condition ${}_1\bar{\theta}_0 + {}_2\bar{\theta}_0 = 3\pi/4$ holds. And the two values of $\bar{\chi}$ are symmetrical about $\pi/2$, as also are those of $\bar{\psi}$. In terms of $\bar{h}$ these values are

$$\bar{\chi}_2 = \sin^{-1} \frac{1}{\bar{h}} \sqrt{\frac{16 - \bar{h}^2}{15}}, \quad \bar{\psi}_2 = \sin^{-1} \sqrt{\frac{16 - \bar{h}^2}{15}}.$$

The upper curve in fig. 8 is drawn from these expressions.

Concluding Remarks.

It is of interest to compare these curves with those first given by Maxwell ('Electricity and Magnetism,' vol. 2) on the postulate (not intended to do more than introduce the phenomenon of hysteresis) that the internal field was expressible by a term, simple harmonic with reference to the displacement, and exhibiting slip when a definite displacement was exceeded. The data regarding the magnetization of crystalline media were not sufficiently extensive then to justify the labour of exact application of Weber's theory. Yet the general features are prominently given. The difference in the results lies mainly in the slight upward curvature of the path where h does not exceed unity, and in the large proportionate value of the residual magnetization where h exceeds unity. If h exceeds 4, the residual magnetization remains constant, and is practically equal to nine-tenths of the saturation magnetization.

When h exceeds 4, all the magnets are held under control. So, if a crystalline material is rotated in a constant field of that magnitude, hysteresis proper ceases. Loss of energy would then appear only through vibrational effects. The smallness of this loss was first observed by Prof. F. G. Baily.

As Maxwell pointed out, the experimentally observed absence of the suddenness of the bend at $h=1$ is explainable by non-uniformity of the molecular magnetic conditions. Fig. 8 is a "reduced" magnetization diagram, applying to any material whatsoever; but the inflexion point will in general differ in different materials because the reducing unit of field strength will in general differ, as also the molecular magnetic moment. It would be easy to draw a graph corresponding to any assumed distribution of inflexion points. The sharp inflexion does not appear even in experiments made on single crystals. This exhibits the non-homogeneity of all crystals. Because the range of the molecular cohesive forces is large in relation to molecular distance, an apparently flawless crystal might exhibit strong magnetic flaws, the law of molecular force being different from that of cohesion. Weber's view leads to the conclusion that a merely incipient flaw extending over only ten molecular distances would constitute a strong magnetic flaw.

It should be noted that the treatment in this note is based on the assumption that the random arrangement of surrounding crystals is such as to give no resultant field on any one crystal. In that case the action is due to the external field alone. But the results do not cease to be of value if the postulate be without foundation. For we would then have to recognize h as being the resultant of the external field and the neglected quasi-external field. Thus, a comparison of the theoretical result with observation might lead to a knowledge of the neglected field.

It is evident that Weber's postulates lead to good elucidation of the normal phenomena of magnetization.

I have to acknowledge with thanks the help given to me by Mr. James Nicoll, M.A., by his accurate calculation, often very heavy, of a great extent of data needed for the whole enquiry, and of which little can appear in this brief synopsis. I also thank the University Court of St. Andrews University for certain annual Grants which enabled me to make use of Mr. Nicoll's services. And I am indebted to Dr. Winifred Smith, member of my departmental staff, for a very careful production of the diagrams.

XVII. The Shot Effect in a Saturated Diode.

By N. R. CAMPBELL, Sc.D., F.Inst.P.,
and

V. J. FRANCIS, B.Sc., A.R.C.S., F.Inst.P. A.M.I.E.E.*

(Communication from the Research Staff of the M.O. Valve Co. Ltd.,
at the G.E.C. Research Laboratories, England.)

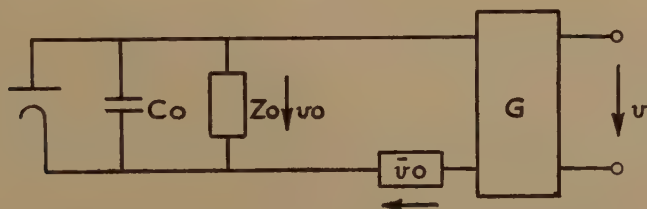
[Received January 27, 1941.]

Summary.

Proves by Heaviside analysis that both Campbell's and Schottky's theorems lead to the same result when applied to calculate the shot noise in the external circuit of a saturated diode, so long as that circuit consists of lumped impedances.

1. With sufficient generality the output circuit of a diode can be taken as composed of the following elements : (1) A condenser C_0 between the

Fig. 1.



electrodes, whose capacity may be merely the inter-electrode capacity but may include that of a condenser connected *directly* across the electrodes ; (2), in parallel with C_0 , a combination Z_0 of impedance elements, whose impedance at pulsance ω is $Z_0(j\omega)$ and through which there is a conductive path ; (3) an amplifier G of infinite input impedance, whose gain at pulsance ω is $G(j\omega)$, and to whose input is applied the difference between v_0 , the potential across Z_0 , and $\bar{v}_0$ the mean value of v_0 . The potential across the output of the amplifier is v . It is assumed that the transit time of the electrons across the diode is infinitely small compared with any other period characteristic of the circuit.

* Communicated by C. C. Paterson.

Owing to the irregularities of electron emission from the cathode, $\overline{v^2}$ will be finite even when the mean current I through the diode is constant. If this current is saturated, two propositions purport to enable $\overline{v^2}$ to be calculated from the circuit constants. One of these is Campbell's theorem, according to which

$$\overline{(v_0 - \bar{v}_0)^2} = I\epsilon \int_0^\infty f^2(t) dt, \quad \dots \dots \dots (1.1)$$

where ϵ is the electronic charge and

$$v_0 = q \cdot f(t) \quad \dots \dots \dots (1.2)$$

represents the variation of v_0 with t after a charge q has been transferred instantaneously between the electrodes at $t=0$. A generalization, due to Rowland*, enables us to find $\overline{v^2}$. We form the Fourier transform of $f(t)$, namely

$$F(j\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} f(t) e^{\pm j\omega t} dt. \quad \dots \dots \dots (1.3)$$

Then
$$\overline{v^2} = 2I\epsilon \int_0^\infty |F(j\omega)G(j\omega)|^2 d\omega. \quad \dots \dots \dots (1.4)$$

The other proposition is due to Schottky. It states that

$$\overline{v^2} = \frac{1}{\pi} I\epsilon \int_0^\infty |Z(j\omega)G(j\omega)|^2 d\omega, \quad \dots \dots \dots (1.5)$$

where Z is the impedance of C_0 and Z_0 in parallel, so that

$$Z(j\omega) = \frac{Z_0(j\omega)}{1 + j\omega C_0 \cdot Z_0(j\omega)}. \quad \dots \dots \dots (1.6)$$

(Some methods of deriving (1.5) seem to suggest that Z should include only impedance elements external to the valve, so that the interelectrode capacity should not be part of C_0 . It is assumed that this suggestion is incorrect.)

(1.4) and (1.5) are based on entirely different physical considerations. But their results will always agree, if

$$2\pi |F(j\omega)|^2 = |Z(j\omega)|^2. \quad \dots \dots \dots (1.7)$$

The purpose of this note is to show that (1.7) is always true, at least if all impedances are lumped.

2. To find $f(t)$ we have to consider the effect of transferring a charge q suddenly from one plate of the condenser, C_0 in the ring of fig. 2, to the other plate. (Since the diode is saturated, its impedance is infinite and can be omitted.) We use the proposition that, so far as the current i

* E. N. Rowland, Proc. Camb. Phil. Soc. xxxii. p. 592 (1936).

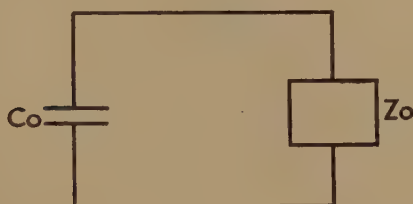
travelling round the ring is concerned, this effect is the same as that of suddenly inserting into the ring a source of constant E.M.F. q/C_0 and of zero impedance. This proposition is well known* and can be used so long as there is not a path through Z_0 composed entirely of condensers in series; it will later be justified for this case also. Then, if $H(t)$ is Heaviside's unit function and p is the Heaviside operator,

$$i = \frac{1}{Z_1(p)} \cdot \frac{q}{C_0} H(t), \quad \dots \dots \dots (2.1)$$

where Z_1 is the impedance of C_0 and Z_0 in series, so that

$$Z_1(p) = \frac{1 + pC_0 \cdot Z_0(p)}{pC_0} \quad \dots \dots \dots (2.2)$$

Fig. 2.



Consequently

$$v_0 = i \cdot Z_0(p) = \frac{q}{C_0} \cdot \frac{Z_0(p)}{Z_1(p)} H(t), \quad \dots \dots \dots (2.3)$$

$$= q \cdot \frac{pZ_0(p)}{1 + C_0 \cdot pZ_0(p)} \quad \dots \dots \dots (2.4)$$

In virtue of the assumption that the impedances are lumped, Z_0 can be expressed as the ratio of two polynomials in p , having no negative powers of p . Further, since there is a conductive path through Z_0 , $Z_0(0)$ is finite. Accordingly, if we write (2.4) in the form

$$v_0 = q \cdot \frac{p\theta(p)}{\phi(p)} H(t),$$

where θ and ϕ are polynomials, the degree of ϕ is not lower than that of $p\theta$ and $p\theta/\phi = 0$ when $p = 0$.

Accordingly †, the solution of (2.4) for $t > 0$ is

$$v_0 = q \sum_a \frac{\alpha \theta(\alpha)}{\alpha \phi'(\alpha)} e^{\alpha t} = q \sum_a \frac{\theta(\alpha)}{\phi(\alpha)} e^{\alpha t}, \quad \dots \dots \dots (2.5)$$

* See, for example, V. Bush, 'Operational Circuit Analysis,' chap. xiv.

† See, for example, H. Jeffreys, 'Operational Methods in Mathematical Physics,' 2nd ed. p. 11, equation (9).

where the α 's are the roots of $\phi(p)=0$, so long as all these roots are different and none is zero. In section 3 we shall proceed on this assumption; in section 4 we shall extend the proof so as to remove the limitation.

3. By definition $v_0/q=f(t)$; and from (1.3) and (2.5)

$$\sqrt{2\pi}F(j\omega)=\int_{-\infty}^{+\infty}\left[\sum_{\alpha}\frac{\theta(\alpha)}{\phi'(\alpha)}e^{(\alpha\pm j\omega)t}\right]dt. \quad (3.1)$$

But, since $v_0=0$ for $t<0$, we may replace the lower limit of integration by 0, and find

$$-\sqrt{2\pi}F(j\omega)=\sum_{\alpha}\frac{\theta(\alpha)}{(\alpha\pm j\omega)\phi'(\alpha)}. \quad (3.2)$$

It is an algebraic identity that, so long as θ is of lower degree than ϕ , which condition is here fulfilled,

$$\sum_{\alpha}\frac{\theta(\alpha)}{(j\omega-\alpha)\phi'(\alpha)}=\frac{\theta(j\omega)}{\phi(j\omega)}. \quad (3.3)$$

But by (1.6)

$$\frac{\theta(j\omega)}{\phi(j\omega)}=Z(j\omega). \quad (3.4)$$

Hence

$$2\pi|F(j\omega)|^2=\left|\sum_{\alpha}\frac{\theta(\alpha)}{(\alpha\pm j\omega)\phi'(\alpha)}\right|^2. \quad (3.5)$$

$$=\left|\sum_{\alpha}\frac{\theta(\alpha)}{(j\omega-\alpha)\phi'(\alpha)}\right|^2. \quad (3.6)$$

$$=|Z(j\omega)|^2. \quad (3.7)$$

and (1.7) is proved.

4. We now remove the restriction on the form of $\phi(p)$. $\phi(p)$ cannot have a zero root, but it may be of the form

$$\phi(p)=(p-\alpha_1)^{r_1}\cdot(p-\alpha_2)^{r_2}\dots(p-\alpha_{\lambda})^{r_{\lambda}}. \quad (4.1)$$

Then, since $\theta(p)$ is of lower degree than $\phi(p)$, we can write, with $m<r_{\lambda}$

$$\frac{p\theta(p)}{\phi(p)}=p\sum_{\lambda}\frac{A_m p^m+A_{m-1}p^{m-1}+\dots+A_0}{(p-\alpha_{\lambda})^{r_{\lambda}}}, \quad (4.2)$$

$$=p\sum_{\lambda}\sum_{n=1}^{r_{\lambda}}\frac{B_n}{(p-\alpha_{\lambda})^n}, \quad (4.3)$$

where the A's and B's are constants independent of p .

By a general proposition of Heaviside analysis

$$\frac{pB_n}{(p-\alpha_{\lambda})^n}H(t)=B_n\frac{t^{n-1}}{(n-1)!}e^{\alpha_{\lambda}t}. \quad (4.4)$$

But

$$\int_0^{\infty}t^{n-1}e^{(\alpha_{\lambda}-j\omega)t}dt=(-1)^n\frac{(n-1)!}{(\alpha_{\lambda}-j\omega)^n}. \quad (4.5)$$

Therefore, using the negative sign in (1.3),

$$\sqrt{2\pi}F(j\omega) = \int_0^\infty \frac{p\theta(p)}{\phi(p)} H(t)e^{-j\omega t} dt. \quad (4.6)$$

$$= \sum_{\lambda} \sum_{n=1}^{\tau_{\lambda}} \frac{B_n}{(j\omega - \alpha_{\lambda})^n}. \quad (4.7)$$

(From (4.3))

$$= \frac{\theta(j\omega)}{\phi(j\omega)} = Z(j\omega). \quad (4.8)$$

5. We now have to justify the procedure of section 2 when there is a path through Z_0 consisting of condensers in series. In this case, the degree of the denominator of $Z_0(p)$ will exceed by 1 the degree of the numerator, and we may write

$$Z_0(p) = \frac{\sum_0^n b_r p^r}{\sum_0^{n+1} a_r p^r}. \quad (5.1)$$

Then

$$Y_0 = \frac{1}{Z_0} = C'p + Y_0', \quad (5.2)$$

where

$$Y_0' = \frac{\sum_0^n a_r' p^r}{\sum_0^n b_r p^r} = \frac{1}{Z_0'} \quad (\text{say}) \quad (5.3)$$

and

$$C'b_{r-1} + a_r' = a_r. \quad (5.4)$$

Z_0 is therefore equivalent for some purposes to Z_0' in parallel with a condenser C' , where Z_0' is a combination through which there is no complete capacitative path. We have to prove it is equivalent for our purpose.

It is clearly equivalent in estimating Z . It will be equivalent so as to justify our assumption if we arrive at the same equation (2.3) whether we take the ring of fig. 2 to be (a) C_0 in series with Z_0 , or (b) a condenser of capacity $(C_0 + C')$ in series with Z_0' .

For (a) we have

$$Z_1 = \frac{1}{C_0 p} + Z_0 = \frac{\sum_0^{n+1} b_r' p^r}{C_0 p \cdot \sum_0^{n+1} a_r p^r}, \quad (5.5)$$

where

$$b_r' = a_r + C_0 b_{r-1}. \quad (5.6)$$

$$i = \frac{q}{C_0} \cdot \frac{Z_0}{Z_1} \cdot H(t) = q \frac{\sum_0^n b_r p^r}{\sum_0^{n+1} b_r' p^r} \cdot H(t). \quad (5.7)$$

For (b) we have

$$Z_1' = \frac{1}{(C_0 + C')p} + Z_0' = \frac{\sum_0^{n+1} b_r'' p^r}{(C_0 + C')p \cdot \sum_0^n a_r' p^r}, \quad (5.8)$$

where

$$b_r'' = (C_0 + C')b_{r-1} + a_r', \quad (5.9)$$

$$i = \frac{q}{C_0 + C'} \cdot \frac{Z_0'}{Z_1'} H(t) = q \cdot \frac{p \sum_0^n b_r p^r}{\sum_0^{n+1} b_r'' p^r}. \quad (5.10)$$

Comparing (5.7) and (5.10), there will be the necessary equivalence if $b_r'' = b_r'$. If in (5.9) we substitute for C' from (5.4), we find that actually $b_r'' = b_r'$.

6. The capacity C_0 does not appear explicitly in (1.7) or (5.10). It might appear therefore that (1.7) would still be true if C_0 were zero. But in assuming that the charge q is transferred *instantaneously* between the electrodes, we assume that during the period of transference, *i. e.*, during the transit of the electrons, no part of the charge induced on the anode by the approach of the charged particle can escape by flowing through the circuit connecting the electrodes. In order that this should be true it is necessary that C_0 should be finite.

XVIII. *On Acoustically Effective Vortex Motion in Gaseous Jets.*

By P. SAVIC, D.Ph. (Basle),

Physics Department, University College, London, at the
University College of North Wales, Bangor *.

[Received April 2, 1941.]

ABSTRACT.

Tollmien's general criterion of instability of one-dimensional fluid motion is applied to the problem of sound sensitive plane jets. Bickley's formula is inserted into Tollmien's disturbance equation and stream function, wave-velocity and wave-length are found. The stream-line representation shows good qualitative agreement with G. B. Brown's photographs of smoke jets. Furthermore, a formula connecting the frequency of the sound with either of the two proper values of the disturbance equation is derived, which is in fair agreement with G. B. Brown's measurements on edge-tones. A brief explanation of the phenomenon of sensitive jets is given.

THE problem of sensitive jets and of edge-tones is still unsolved. Although Kruger and Schmidtke ⁽¹⁾ provided a theory of edge-tones, Brown ⁽²⁾ has shown that their conclusions are only approximately valid. Brown suggested that the treatment which hitherto only considered an ideal fluid should be extended to viscous and compressible fluids. Such a treatment may then be capable of explaining both the phenomena of edge-tones and of sensitive jets. Richardson ⁽³⁾, who provided a different theory of edge-tones which overcame some of the difficulties encountered by Kruger's considerations, equally assumes an ideal fluid, but was unable to explain all phenomena observed, especially the apparently closely-related feature of sensitive jets.

Since Brown ⁽⁴⁾, in his research on sensitive jets, found no agreement of his experiments with Rayleigh's considerations on vortex motion in an ideal fluid, it seems probable that the fundamental constants of a real fluid, viz., viscosity and compressibility, cannot be neglected. The mathematical difficulties connected with such general considerations of a real fluid seem, however, insurmountable. The present paper will show

* Communicated by the Author.

that the introduction of small viscosity alone is possible, and that it results in a considerable improvement of the agreement with experiment.

The problem of turbulence has always been connected with the necessity of introducing viscosity. Tollmien ⁽⁵⁾ succeeded recently in deriving a general theorem of turbulence in a viscous fluid. He proceeded as follows:—The fluid motion was considered to be two-dimensional, and the main velocity distribution was assumed to depend only on the y -direction perpendicular to the main flow; thus $U=U(y)$. Superposing small velocity variations, introducing them into the Stokes-Navier equations, eliminating the pressure and neglecting terms of second order led to a linear differential equation. The linearity suggested the building-up of the disturbance stream function Ψ from partial vibrations, which were assumed to be of the form :

$$\Psi = \phi(y) \cdot \exp. i\alpha(x-ct), \quad \dots \dots \dots (1)$$

where α is a real constant, but c is in general complex. The wave-length λ of the disturbance is $\lambda=2\pi/\alpha$, the real part of c is the wave-velocity. This assumption finally leads to the general equation governing the development of the disturbance

$$(U-c)(\phi''-\alpha^2\phi)-U''\phi=-\frac{i}{\alpha R}(\phi'''-2\alpha^2\phi''+\alpha^4\phi), \quad \dots \dots (2)$$

where a non-dimensional system of units is employed by the introduction of a representative length and a representative velocity. The viscosity thus appears in the Reynolds number R . For an inviscid fluid ($R=\infty$) the equation (2) reduces to the simpler one :

$$(U-c)(\phi''-\alpha^2\phi)-U''\phi=0. \quad \dots \dots \dots (3)$$

This equation has a singularity at $U=c$. Tollmien has shown that this singular point cannot be removed unless some of the viscosity terms are retained. The removal then leads to a general criterion for turbulence: All profiles with a point of inflexion are unstable and tend to exhibit oscillations. Both increased and neutral oscillations (c entirely real) exist, and for the neutral ones the wave velocity is equal to the stream velocity at the point of inflexion of the velocity profile. Although these results are derived for the flow between solid walls, Hollingdale ⁽⁶⁾ has shown that they also hold for an unlimited fluid.

Two experimental researches testing this theory have so far been carried out, one by Rosenbrook ⁽⁷⁾, the other by Hollingdale. Rosenbrook examined the flow in a slightly diverging channel; Hollingdale investigated the motion in the wake of a solid body. Both authors approximated the observed velocity profile by a mathematical expression which was introduced into the disturbance equation (3). Both achieved a fair agreement with the theory as regards the proper values c and α .

The question now arises whether the theory still holds if the angle between the solid walls, which in Rosenbrook's research was taken to be 6° , is increased until it reaches the value 180° . This would mean that the fluid emerges from a slit into an unlimited sea. For the case of a linear slit Bickley ⁽⁸⁾ has derived a formula which has been verified by Andrade ⁽⁹⁾. But here the dependence of the velocity upon the x -direction parallel to the main flow cannot be neglected. This is probably the reason why the production of jet-tones is far more difficult and the tone itself far less clear than in the case of Aeolian tones, which are produced in the wake of a solid body. The uniformity of the velocity distribution in the x -direction in such a wake is in fact better, as Hollingdale indicated, although the slight drop along the x -axis has to be taken into account in order to determine the size of the velocity profile at the point on the x -axis, where instability first sets in. Nevertheless, the existence of jet-tones suggests that the Tollmien-Hollingdale theory can be applied to the linear jet too.

Bickley's velocity profile is given by the expression

$$U = a \cdot \text{sech}^2 by, \quad (4)$$

where

$$a = 0.454 \dots (K^2/\nu x)^{1/3}$$

and

$$b = 0.2751 \dots (K/\nu^2)^{1/3} \cdot x^{-2/3}. \quad (5)$$

ν is the kinematic viscosity and K the momentum of the fluid per unit time and unit length of the slit divided by the density. Twofold differentiation shows that two points of inflexion exist at

$$U_i = c = 2a/3. \quad (6)$$

Inserting (4) and (6) into the disturbance equation (3) we obtain

$$\phi'' - \alpha^2 \phi + 6b^2 \phi \cdot \text{sech}^2 by = 0. \quad (7)$$

The boundary conditions for ϕ are :

$$\begin{cases} \phi \text{ an even function of } y \text{ (in order to obtain an alternate vortex street),} \\ \phi = 0 \text{ as } y = \infty. \end{cases}$$

The neutral solution of equation (7), which also satisfies the required boundary conditions, is

$$\phi = k \cdot a \cdot \text{sech}^2 by \quad (8)$$

(k is an arbitrary constant) and the proper value becomes

$$\alpha = 2b \quad \text{or} \quad \lambda = \pi/b. \quad (9)$$

The complete instantaneous stream-function is, from (1),

$$\Psi_1 = \int_0^y U \cdot dy + ka \cdot \text{sech}^2 by \cos 2bx = \frac{a}{b} \tanh by + ka \cdot \text{sech}^2 by \cdot \cos 2bx, \quad (10)$$

where the real part has been taken.

It is interesting to compare this result with Karman's vortex street. It can be seen that the stream-lines are symmetrical to the lines $\cos 2bx=1$. The centres of the vortices must therefore lie on these lines. Assuming this, we can write (10) in the form

$$\tanh^2 by - \frac{1}{bk} \tanh by + \frac{\Psi_1}{a.k} - 1 = 0. \quad (11)$$

For the centres of the vortices the discriminant of this equation must vanish. This is fulfilled if

$$\Psi_1 = \frac{a + 4b^2 k^2 a}{4b^2 k} \quad \text{and} \quad \tanh by = \frac{1}{2bk}.$$

Calling the distance between the two rows of vortex centres h , we obtain

$$h = \frac{2}{b} \cdot \tanh^{-1} \frac{1}{2bk},$$

and from (9)

$$\frac{h}{\lambda} = \frac{2}{\pi} \tanh^{-1} \frac{1}{2bk}.$$

Karman finds

$$\frac{h}{\lambda} = \frac{1}{\pi} \tanh^{-1} \frac{1}{\sqrt{2}}.$$

The two formulæ agree only if

$$bk = \frac{1}{\sqrt{2}} \pm \frac{1}{2}.$$

Since k is an arbitrary constant it is clear that no value for h/λ can be derived from this theory.

In order to compare our stream-function with Brown's photographs of sensitive jets, we must bear in mind that these photographs represent the paths of smoke particles at a particular moment or for an observer moving with the wave-velocity c in the x -direction at any moment. We therefore have to subtract the stream-function of the velocity c from formula (10), thus :

$$\Psi_2 = -\frac{2}{3}ay + \frac{a}{b} \tanh by + ak \cdot \text{sech}^2 by \cdot \cos 2bx.$$

The figures 1 and 2 show two such stream-line fields for $k=1$ and $k=0.1$. Fig. 2 is in fair qualitative agreement with Brown's photographs, especially with those with small increment, *e. g.*, fig. 5 of reference (4). Fig. 1 illustrates the state of motion in a water-jet, where the amplitude is larger than in gases. b was taken to be equal to $1/\sqrt{2}$.

We can now proceed to find an expression for the relative wave or vortex velocity c/v , where v is the stream-velocity at the slit orifice. Andrade has indicated on what conditions this velocity can be assumed

to be uniform across the slit. These conditions are, in fact, satisfied in Brown's experiments. We can therefore express the momentum by

$$K = m.v^2, \quad (12)$$

where m is the width of the slit. Eliminating x between a and b in (5)

Fig. 1.

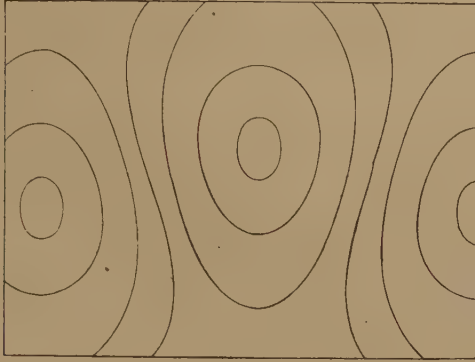
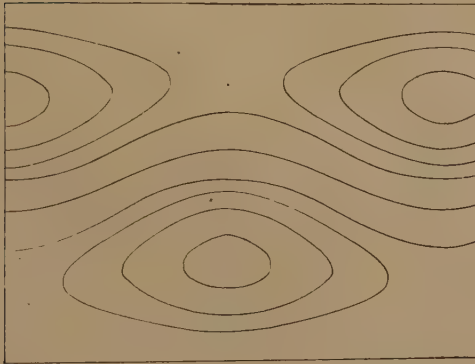


Fig. 2.



and inserting (6) and (9) we obtain, after making use of (12) and the relation $\lambda = c/n$ (n -frequency), the final result

$$\lambda = \left(\frac{m.v^2}{0.9541.n^2} \right)^{1/3} \quad (13)$$

and

$$\frac{c}{v} = \left(\frac{m.n}{0.9541.v} \right)^{1/3} \quad (14)$$

These formulæ can be tested by Brown's experimental results. In his paper on sensitive jets [reference (4)] Brown has defined the wave-length as the distance between two successive vortices, the centres of which have

been estimated from the photographs. Because this distance largely depends on the increment of the oscillations, the agreement with the above formulæ cannot be expected to be satisfactory. Re-measuring his films with the different convention, that the wave-length shall be defined as the distance between equal phases of the sinusoidal motion, Brown found, however, that there is no difference between the vortex motion in sensitive jets and in edge-tone jets, provided that the stream-velocity v , the slit-width m and the frequency n is the same in both cases. We can therefore confine our attention to the motion in edge-tone columns. From the Tables II. and III. of reference (2) the values of n , U , and m have

TABLE I.

n .	v cm./sec.	c/v theor.	c/v exp.	λ theor.	λ exp.
97	215	0.36	0.38	0.80	0.85
97	248	0.35	0.36	0.88	0.93
194	321	0.40	0.41	0.66	0.68
97	210	0.36	0.37	0.79	0.81
194	380	0.38	0.39	0.74	0.77
291	391	0.43	0.42	0.57	0.56
388	508	0.43	0.43	0.56	0.57
97	227	0.36	0.38	0.83	0.89
291	422	0.42	0.39	0.60	0.57
388	550	0.42	0.42	0.60	0.59
97	200	0.37	0.38	0.76	0.78
194	345	0.36	0.36	0.69	0.64
291	502	0.39	0.37	0.68	0.64
97	268	0.34	0.36	0.93	1.00
291	342	0.45	0.38	0.52	0.45
485	517	0.46	0.40	0.49	0.43
800	616	0.51	0.47	0.40	0.36

$$m=0.1 \text{ cm.}$$

been inserted into our formulæ (13) and (14). U in these tables has been termed v and u/U in Brown's paper is equal to c/v in our formula. The comparison of the theoretical values of λ and c/v with Brown's experimental ones are shown in our Tables I. and II. The agreement is fairly good, considering the degree of approximation in neglecting the variation of the velocity in the x -direction and taking the neutral type of oscillations instead of the augmenting one.

An interesting feature is the bad agreement of the last four lines. Here the height of the edge above the orifice is very small. Since Bickley's formula does not hold in the region close to the orifice the agreement of the

experiment with our formulæ must be bad too. A similar argument holds for the case of small stream velocity v .

The mechanism of sensitive gaseous jets can now be explained in a few words. The jet is sensitive to sound, because its velocity profile has a point of inflexion. The reason why the jet can easily be controlled by an outside sound lies in the fact that the size of the velocity profile U and therefore its appropriate frequency changes with x . There is therefore no reason why the undisturbed jet should break down to instability at a particular point and thus emit a particular frequency.

TABLE II.

n .	v cm./sec.	c/v theor.	c/v exp.	λ theor.	λ exp.
20	138	0.25	0.22	1.71	1.51
21	142	0.25	0.23	1.69	1.56
25	162	0.25	0.23	1.64	1.51
29.5	187	0.25	0.24	1.61	1.50
36	200	0.27	0.28	1.48	1.55
36	140	0.30	0.26	1.17	1.00
50	185	0.30	0.28	1.13	1.05
53	186	0.31	0.29	1.09	1.02
58.8	110	0.31	0.29	1.10	1.04
59	212	0.31	0.30	1.11	1.07
71.5	245	0.31	0.31	1.07	1.08
75	254	0.31	0.30	1.06	1.02
84	277	0.32	0.34	1.05	1.11
84	271	0.32	0.31	1.03	1.00
89	283	0.32	0.32	1.02	1.02
125	283	0.36	0.40	0.81	0.90
126	272	0.37	0.40	0.79	0.87
143	305	0.37	0.41	0.78	0.88
147	311	0.37	0.38	0.78	0.81

The outside sound determines this point and makes the jet oscillate as if it would emit a jet-tone with the same frequency. Although this treatment provides a solution which agrees better with experiment, Rayleigh's theory of an ideal fluid is still unable to explain all the facts observed, namely the occurrence of a large number of maxima and minima of sensitivity of the jet, which are only dependent upon the frequency. This feature, which seems to lie completely beyond the realm of kinematic similarity, is very probably due to the influence of the compressibility of the outside air, as Brown suggested.

I am indebted to Mr. D. S. Box for a valuable suggestion.

References.

- (1) Kruger and Schmidtke, *Ann. d. Physik*, lx. p. 701 (1919).
- (2) Brown, G. B., *Proc. Phys. Soc.* xlix. p. 508 (1937) and p. 493 (1937).
- (3) Richardson, E. G., *Proc. Phys. Soc.* xliii. p. 239 (1931).
- (4) Brown, G. B., *Proc. Phys. Soc.* xlvii. p. 703 (1935).
- (5) Tollmien, W., *Nachr. Ges. Wiss. Gottingen, Math.-Phys. Klasse*, p. 79 (1935).
- (6) Hollingdale, S. H., *Phil. Mag.* xxix. p. 209 (1940).
- (7) Rosenbrock, G., *Zeitschr. angew. Math. Mech.* xvii. p. 8 (1937).
- (8) Bickley, W., *Phil. Mag.* xxiii. p. 727 (1937).
- (9) Andrade, E. N. da C., *Proc. Phys. Soc.* li. p. 784 (1939).
- (10) v. Karman, Th., and Rubach, H., *Phys. Zeitschr.* xiii. p. 48 (1912).

XIX. *On a Relation between Viscosity and Density of Liquids and Liquid Mixtures.*

By M. K. SRINIVASAN *.

[Received May 27, 1940.]

GOOD accounts of the work on the viscosity of liquids as a function of temperature, free space, and density are given by Ferguson in the 'Reports on Progress in Physics' (vol. xi. 1935), and in the 'First and Second Reports on Viscosity and Plasticity' ⁽¹⁾. The equations of Batschinski ⁽²⁾, Mott Saunders ⁽³⁾, and Newton Friend ⁽⁴⁾ refer to viscosity in terms of density, and are applicable to unassociated liquids over short ranges of temperature.

The author has shown that an equation of the type $\log \eta = \alpha + \frac{\beta}{\rho_0 - \rho}$, where η is viscosity, ρ density, α , β , and ρ_0 are constants, works well. The equation is shown to apply for unassociated and associated liquids over wide ranges of temperature. It holds even for those associated liquids like the tertiary alcohols, whose association changes with temperature. The only case where it fails (among the liquids studied) is that of water.

Since density is a simple function of temperature for short ranges ($\rho = \rho_0 - bT$), one may obtain the above equation by substituting $\frac{\rho_0 - \rho}{b}$ for T in Andrade's equations ^{(5) (6)}; $\eta = Ae^{b/T}$ or $\eta = A'e^{b/T - \theta}$. The equation holds for cases in which ρ is not rigidly linear with temperature because of some compensation resulting from the constants.

Tables I. to III. give in detail the results for octane, 3-methyl heptanol 3, and gallium. The other results are summarized in Table IV., where the constants α , β , and ρ_0 are given along with the maximum percentage difference and range $\left(\frac{\eta_{\text{maximum}}}{\eta_{\text{minimum}}} \right)$. The last two columns show for comparison the range and maximum percentage difference in the case of Andrade's modified equation ⁽⁷⁾; $\eta v^{\frac{1}{3}} = Ae^{c/vT}$. The variation of viscosity of binary mixtures with temperature (composition remaining the same) is also represented by the equation. The results are summarized in Table V.

The data have been obtained from Landolt-Bornstein Tabellen, *Tables Annuelles*, or from sources indicated in the Tables.

* Communicated by Professor B. Prasad.

TABLE I.

Octane, η range 3.27; $\alpha=4.1439$, $\beta=0.4000$, $\rho_0=0.9534$.

Temperature (°C.).	ρ .	$\eta \times 10^6$ (obs.).	$\eta \times 10^6$ (cal.).	Percentage difference.
0	0.7185	7060	7028	0.4
20	.7022	5419	5449	0.6
40	.6860	4328	4362	0.7
50	.6778	3907	3939	0.8
60	.6694	3551	3567	0.6
70	.6611	3241	3254	0.4
80	.6525	2971	2971	0.0
90	.6438	2730	2728	0.1
100	.6351	2520	2516	0.2
110	.6260	2335	2321	0.8
120	.6168	2160	2149	0.5

TABLE II.

3-methyl heptanol 3, η range 34.9; $\alpha=4.5172$, $\beta=0.19716$, $\rho_0=0.91615$.

Temperature (°C.).	ρ .	$\eta \times 10^6$ (obs.).	$\eta \times 10^6$ (cal.).	Percentage difference.
0	0.84538	198690	201000	1.1
10	.83712	106060	102800	3.1
20	.82924	61854	61066	1.3
30	.82105	38911	38941	0.1
40	.81259	26056	26363	1.1
60	.79458	13563	13772	1.6
80	.77605	8292	8404	1.4
100	.75752	5691	5756	1.1

TABLE III.

Gallium ⁽⁸⁾, η range 3.27, $\alpha=3.4591$, $\beta=0.3004$, $\rho_0=6.444$.

Temperature (°C.).	ρ .	$\eta \times 10^6$ (obs.).	$\eta \times 10^6$ (cal.).	Percentage difference.
52.9	6.080	18940	19240	1.6
102	6.041	16040	16020	0.1
149	6.005	14060	13920	1.0
200	5.972	12580	12470	0.9
301	5.905	10290	10390	1.0
402	5.840	8820	9048	2.6
500	5.779	8127	8143	0.2
600	5.720	7637	7485	2.0
806	5.604	6524	6558	0.5
1010	5.492	5915	5953	0.6
1100	5.445	5783	5751	0.6

TABLE IV.

Liquid.	Temp. range (°C.).	a.	β .	ρ_0 .	η range.	Max. % difference.	Andrade:	
							η range.	Max. % difference.
Pentane	0-30	4.5114	0.19187	0.8474	1.32	0.2	1.36	0.4
Hexane	0-60	4.4497	0.23458	0.8800	1.76	0.6	1.93	0.14
Heptane	0-90	4.2955	0.31600	0.9221	2.34	0.6	2.29	0.8
Octane	0-120	4.1439	0.40000	0.9534	3.27	0.8	3.45	0.8
Benzene	0-80	4.7090	0.20000	1.060	2.70	0.9	2.37	0.8
Methyl alcohol	0-60	5.4544	0.76719	1.122	2.29	1.2	2.27	0.6
Ethyl alcohol	0-70	5.5086	0.68675	1.057	3.48	0.5	3.22	1.0
Propyl alcohol	0-90	5.9031	0.43621	0.9817	7.31	0.7	6.61	2.6
(9) 4-methyl heptano 1-3	0-100	4.5375	0.21642	0.9385	5.07	1.7	5.07	4.0
(9) 4-methyl heptano 1-4	0-100	4.4494	0.19534	0.9036	59.6	5.0	59.6	25
(9) 3-methyl heptano 1-3	0-100	4.5172	0.19716	0.91615	34.9	3.1	34.9	20
Triethyl carbinol	0-100	4.6122	0.16012	0.91475	68.6	4.8	68.6	30
Methyl acetate.	0-50	4.5000	0.34285	1.2486	1.65	0.1	1.75	0.1
Ethyl acetate	0-70	4.5213	0.31189	1.175	2.09	0.2	2.23	0.4
Propyl acetate	0-100	4.2444	0.45878	1.189	2.99	1.2	2.96	0.6
Propyl formate	0-80	4.4600	0.35286	1.187	2.33	0.4	2.33	0.2
Amyl butyrate.	0-100	4.4710	0.31821	1.062	3.98	1.2	3.98	2.8
Carbon tetrachloride.	0-110	4.6259	0.55438	2.000	2.53	1.0	2.66	1.1
Acetic acid	0-110	4.5403	0.36543	1.2854	2.88	2.0	2.54	0.3
(10) Tin	240-800	3.4094	0.46677	7.518	1.67	0.8	1.67	2.1
(11) Sodium	98-355	4.99681	0.075456	1.0890	2.41	2.2	2.41	4.2
(11) Potassium	64-352	3.02332	0.06172	1.0974	2.25	3.4	2.25	3.6
Gallium	53-1100	3.4591	0.3004	6.444	3.27	2.6	3.27	3.9

TABLE V.

System, Benzene Benzyl benzoate ⁽¹²⁾. Temperature range 5 to 60° C.

Per cent. benzene.	α .	β .	ρ_0 .	Maximum per cent. difference.
0	4.8232	0.25363	1.237	1.2
25	4.8234	0.24150	1.191	2.3
50	4.5631	0.29800	1.167	0.8
75	4.4219	0.34403	1.150	0.4
100	4.4220	0.31434	1.104	0.3

TABLE VI.

System, Ethyl ether and Phenetol ⁽¹⁵⁾; $\alpha=8.2630$, $\beta=10.219$, $\rho_0=2.7241$.

Mol. fraction of ethyl ether.	Density.	η (obs.).	η (cal.).	Per cent. difference.	Per cent. difference (Macleod).
1.0000	0.7139	0.002233	0.002221	0.5	0
0.9004	.7458	0.002674	0.002679	0.2	1.9
0.8303	.7656	0.003057	0.003027	1.0	1.2
0.7485	.7891	0.003494	0.003499	0.2	2.1
0.6882	.8059	0.003866	0.003890	0.9	2.5
0.6048	.8282	0.004406	0.004498	2.1	2.1
0.5579	.8387	0.004809	0.004819	0.2	0.7
0.5106	.8488	0.005154	0.005155	0.0	0.1
0.4501	.8628	0.005713	0.005666	0.7	2.2
0.3516	.8853	0.006667	0.006613	0.8	2.0
0.3026	.8970	0.007152	0.007178	0.4	1.4
0.2523	.9083	0.007856	0.007780	1.0	0.9
0.1855	.9230	0.008671	0.008648	0.3	1.9
0.0975	.9427	0.009993	0.009986	0.1	0.8
0.0745	.9471	0.01027	0.01032	0.5	0.0
0.0000	.9618	0.01158	0.01153	0.5	0

TABLE VII.

System.	α .	β .	σ_0 .	Maximum per cent. difference.	
				(Author).	(Macleod).
(15) Diethyl ether and diphenyl ether	6.4470	3.5278	1.9240	3.3	3.8
(15) Ethyl acetate and ethyl benzoate	4.4264	0.4825	1.2974	4.9	4.2
(15) Ethyl benzoate and benzyl benzoate	6.4906	1.8789	1.5365	3.1	3.6
(15) Ethyl acetate and benzyl benzoate	5.8960	0.89558	1.4080	5.0	3.0
(15) Phenetol and diphenyl ether	5.7470	1.3500	1.5454	1.6	2.4
(14) Phenol and benzene.	5.7024	1.1154	1.4090	3.5	8.6
(15) Toluene and benzyl benzoate	5.5360	1.5787	1.5778	4.5	—

As may be seen from the tables the equation compares well with Andrade's equation. Particular mention may be made of the fit obtained with the tertiary alcohols for which Andrade's equation does not give good results. The list of substances studied does not include a sufficient number of homologous compounds to look for any generalizable regularity among the constants in each series.

A special feature of this equation is its applicability to ideal binary liquid systems at constant temperature at various compositions. While the variation of viscosity of a binary mixture with temperature (at constant composition) can be represented by Andrade's equation, it has not been possible so far to represent the variation of η of an ideal mixture at different compositions by an equation of the Andrade type. Spells⁽¹³⁾ has proposed an equation (based on Andrade's equation) which works well for non-ideal binary mixtures only. In the following tables results are shown for a few ideal binary liquid mixtures. In the last column percentage difference in the case of Macleod's equation⁽¹⁴⁾:

$$\eta = \eta_1 m_1 \frac{x_1}{x} + \eta_2 m_2 \frac{x_2}{x}$$

(where η_1 and η_2 are the viscosities of the components, m_1 and m_2 their molecular percentages, x_1 and x_2 their amounts of free space per cubic centimetre, and x the free space of the mixture) is reproduced from his paper.

In conclusion, I wish to thank Professor B. Prasad for his kind encouragement and interest in this study. I am also grateful to the Government of Orissa for the grant of a research scholarship.

Summary.

(1) An equation of the type $\log \eta = \alpha + \beta/(\rho_0 - \rho)$ (where η is viscosity, ρ density, and α , β , and ρ_0 are constants) is proposed and is shown to represent the variation of viscosity of unassociated and associated, organic and inorganic liquids, with temperature.

(2) The same equation represents the viscosity variation with temperature of binary liquid mixtures at constant composition.

(3) It also represents at constant temperature the viscosity variation of ideal binary liquid mixtures with composition.

References.

- (1) 'Reports on Viscosity and Plasticity,' Academy of Sciences, Amsterdam ; First Report, 1935 ; Second Report, 1939.
- (2) Batschinski, *Z. Physik. Chem.* lxxxiv. p. 696 (1913).
- (3) Mott Sayers (Jr.), *J. Amer. Chem. Soc.* lx. p. 154 (1938).
- (4) Newton Friend, *Trans. Faraday Soc.* xxxiv. p. 813 (1938).
- (5) Andrade, 'Nature' (March 1, 1930).
- (6) Andrade, 'Nature' (April 12, 1930).
- (7) Andrade, *Phil. Mag.* (7) xvii. p. 497 (1934) ; *ibid.* (7) xvii. p. 698 (1934).
- (8) Spells, *Proc. Phys. Soc.* xlviii. p. 299 (1936).
- (9) Bingham and Darrall, *J. Rheology*, i. p. 174 (1930).
- (10) Stott, *Proc. Phys. Soc.* xlv. p. 530 (1933).
- (11) Chiong, *Proc. Roy. Soc. A*, clvii. A, p. 264 (1936) ; "Density from Relative Values of Specific Volume given by Macleod," *Proc. Phys. Soc.* i. p. 788 (1938).
- (12) Bingham and Brown, *J. Rheology*, iii. p. 95 (1932).
- (13) Spells, *Trans. Farad. Soc.* xxxii. p. 530 (1936).
- (14) Macleod, *Trans. Farad. Soc.* xx. p. 348 (1924).
- (15) Kendall and Wright, *J. Amer. Chem. Soc.* xlii. p. 1776 (1920).
- (16) Kendall and Munroe, *J. Amer. Chem. Soc.* xxxix. p. 1787 (1917).

Mayurbunj Chemical Laboratory,
Ravenshaw College,
Cuttack, India.

the leading note in the key of *D*, is a comma higher than *C*, a major third above *A*.

4. The adoption of the comma as an integer introduces a new musical scale and, although a gamut with 54 notes to the octave may appear to be unsuitable for a manual keyboard, the development of mechanically played music must not be ignored.

The King's Lodge,
Hunton Bridge, Hertfordshire.

XXI. *Notices respecting New Books.*

Geomagnetism. By SYDNEY CHAPMAN and JULIUS BARTELS. Two volumes. [Vol. I., pp. xxviii+1 to 542; Vol. II. pp. x+543 to 1049.] (Oxford: At the Clarendon Press, 1940. Price 63s. net.)

THIS is a great work—great in conception, great in execution, and great in weight; its two volumes turn the scale at about eight-and-a-half pounds. A dozen years (and who shall say how many years of investigation?) have gone to the making of the book, which is the only modern survey of its kind and is likely to remain the standard treatise on the subject for many years to come.

The first volume, which deals with the observed phenomena of geomagnetism, is mainly factual, and those who, with the present writer, love an instrument, will find much to interest them in the descriptions of experimental methods. The authors are not afraid to begin at the very beginning of the subject and have raised a logical structure for which all subsequent writers on magnetism will be grateful. It is to be hoped that they will not forget to express their gratitude.

The second volume is divided into two parts; the first part deals with the analysis and synthesis of geomagnetic data and the second with physical theories of geomagnetic phenomena. The volume closes with a valuable series of historical notes, a comprehensive bibliography of some twelve hundred references, a series of tables and a very detailed index.

Errors of detail there may be, but they are few, and it would be impertinent, both in the older and later sense of the word, to attempt to discuss them in a necessarily brief review of a work so wide in its scope and so generous in its scholarship. The treatise will receive a warm welcome from all serious students of terrestrial magnetism. To such students it is quite indispensable.

The Modern Theory of Solids. By Dr. F. SEITZ. [Pp. xv+698.] (London: McGraw-Hill Publishing Co., Ltd., 1940. Price 49s.)

THIS is one of the most important of the books recently issued by the publishing house of McGraw-Hill. The author, in his very wide and scholarly review, has attempted to reach three classes of audience—the student without any specialized knowledge of this rapidly growing branch of physics; the experimenter—physicist, chemist, metallurgist, or engineer—who may be “keeping an eye on a field of physics that is of possible value” to him; and the theoretical physicist. It is no easy task to write up (or down) to so varied an audience, but the author has succeeded. The more difficult sections are asterisked and the reader of limited mathematical attainments may concentrate on the conclusions of these sections, rather than on the arguments by which these conclusions are reached.

The book covers a wide field. The classical theory of ionic crystals; the free-electron theory of metals; the many-body problem; the Band approximation; cohesive energy; the work function; the excited electronic states; the electronic structure of the five solid types; the dynamics of nuclear motion; the theory of conductivity; magnetic and optical properties. These are the main topics handled in this very comprehensive survey.

The book is clearly written, but—and this is not said by way of criticism—it is not easy reading, and demands the closest attention on the reader's part. He will be well rewarded for his trouble.

The Separation of Gases. By Dr. M. RUHEMANN. [Pp. xiii+283.] (Oxford: The Clarendon Press, 1940. Price 21s. net.)

THE study of what the author terms "deep refrigeration" has received much attention in the Soviet Union, and two Russian treatises on the subject have recently appeared from the pens of Prof. Gersh and Prof. Pavlov.

The author feels that the subject has not hitherto received adequate attention in this country, and his pages have been written to the end that interest may be evoked in what is undoubtedly an attractive and important subject. Rather less than half the book is devoted to general thermodynamical considerations and experimental methods; four chapters are given to the separation of air; the concluding chapters deal with the separation of coke-oven gas, methane and helium, and the separation of olefines from cracker-gas.

The book is a refreshing study, from the academic angle, of what are large-scale engineering problems, and the author is to be congratulated on his presentation of a difficult and fascinating subject. A. F.

Chemical Thermodynamics. By Prof. J. R. PARTINGTON. [Pp. x+230.] (London: Constable & Co., 1940. Price 14s. net.)

THIS, the third edition of Professor Partington's well-known text-book, has been completely re-set, and the many alterations and additions make it virtually a new book. Nevertheless it still remains modest in style and price, and retains the excellent qualities of its predecessors. There is much slipshod thinking in many discussions of thermodynamics as applied to physical chemistry, and students trained in Professor Partington's methods of presentation of the subject are less likely than most to be entrapped by the pitfalls that await the unwary traveller in these domains. This edition contains references to many recent advances and includes a new chapter on statistical mechanics.

It may be recommended as an introduction which will lead up to more detailed studies, and will leave the student with nothing to unlearn. A. F.

The Calculus of Extension. By HENRY GEORGE FORDER. [Pp. xv+490.] (Cambridge University Press, 1941. Price 30s. net.)

THE subject of this book is Grassmann's space analysis, as seen through modern eyes. Definitions of the elements to be operated on, and specifications of the processes and operations are laid down abstractly, thus developing an algebra, which is only subsequently interpreted in geometrical terms. This is not to say that the whole of the calculus is developed before any of the interpretation is given, but only that at each step the formal properties are given before the interpretation.

Grassmann's concepts, inner and outer products and the like, can of course be interpreted geometrically in different ways, and in the early part of the book the interpretation is confined to the case where the operands are geometrical

points and vectors. In the later parts of the book the advanced theory is set out in a most useful way, the chapters on general theory of matrices and on algebraic products being particularly stimulating. The author's personal interest is clearly in the geometrical applications of the method, and he gives many detailed theorems on the properties of conics and the geometry of the triangle, which certainly demonstrate the power of these methods.

Applied X-Rays. By Professor G. L. CLARK. [Pp. xvii+674.] (London: McGraw-Hill Publishing Company, Ltd., 1940. Price 42s.)

THE first edition of Prof. Clark's treatise was published in 1927, the second in 1932. In the eight years that have elapsed the application of X-rays have been so many and so important, not to say spectacular, that it must have cost the author much heartsearching to decide on what to chronicle and what to omit. The author remarks that in the field of cellulose alone, available reprints of papers since 1932 make a pile nearly two feet high. This edition has been completely recast, and becomes practically a new book. It is no mere collection of abstracts, but a co-ordinated whole, which represents the results of a dozen years' teaching of the subject to graduate students.

The book is divided into two parts. Part I. (General Principles and Applications) is concerned with historical matter, equipment, measurement of intensity and quality, atomic structure, chemical analysis, photochemistry, absorption and scattering, radiography and biological applications. Part II. is entirely devoted to the X-ray analysis of the ultimate structures of materials, and covers that very wide field in what is necessarily a condensed but masterly survey.

The book may be read with interest and profit by experts, and by those physicists who, not being specialists in any of these fields, would know something of the very remarkable advances made during the last decade in a rapidly developing subject.

[The Editors do not hold themselves responsible for the views expressed by their correspondents.]

XXII. *Non-linear Distortion, with particular reference to the Theory of Frequency Modulated Waves.*—Part I.

By E. C. CHERRY, M.Sc., and R. S. RIVLIN, M.A.

(Communication from the Staff of the Research Laboratories of The General Electric Company Limited, England)*.

[Received April 21, 1941.]

SUMMARY.

Explicit expressions are derived for harmonic and cross-modulation products in the output current of a non-linear resistive device, such as a thermionic valve, resulting from the application of a wave of E.M.F. This wave may consist of any number of sinusoidal components.

The sideband amplitudes of a carrier wave, frequency modulated by any number of sinusoidal components, are derived and the output current resulting from the application of such a wave to the non-linear device is analyzed.

1. *Introduction.*

IF the characteristic of a thermionic valve is linear, then the output current is directly proportional to the applied E.M.F. except for a constant current. If the valve characteristic is non-linear, then this proportionality will no longer be true, and a sinusoidally varying applied E.M.F. of angular frequency ω (say) will no longer give a sinusoidally varying output current, but will, in general, give a current containing components of angular frequency ω , 2ω , 3ω , $\dots r\omega$, $\dots$, etc.

Again, if two or more signals of different angular frequencies $\omega_1, \omega_2, \dots \omega_n, \dots$ are applied simultaneously to the valve, the output current will generally contain components whose angular frequencies are $(p\omega_1 + q\omega_2 + \dots + r\omega_n + \dots)$, where $p, q, \dots r, \dots$, etc., can take positive or negative integral values or the value zero.

The problem of determining, in these two cases, the amplitude and phase of each component in the output current for a valve having a given non-linear characteristic has been considered by many writers. Thus, Bartlett^{(1), (2)} has taken a non-linear characteristic given by the equation

$$i = A_0 + A_1 v + A_2 v^2 + \dots + A_n v^n + \dots, \quad (1)$$

* Communicated by Dr. C. C. Paterson.

where the coefficients $A_0, A_1, A_2, \dots, A_n, \dots$ are assumed known, while i and v are the output current and applied E.M.F. respectively, and he has calculated the modulation products resulting from the application of certain combinations of sinusoidal waves. Espley⁽³⁾ has discussed the method of calculating the coefficients in equation (1) from experimentally determined points, at equally spaced intervals of v on the valve characteristic, and has evaluated the various harmonics present in the output current when the applied signal is purely sinusoidal. In a subsequent paper⁽⁴⁾, he has shown how the work of Bartlett^{(1), (2)} may be applied in practical cases. Espley⁽⁵⁾ has developed his method of calculation of harmonic content of the output current for a purely sinusoidal input signal, using the Newton-Stirling central difference formula, applied to points on the valve characteristic at equally spaced intervals of v , to determine the coefficients of the polynomial in the equation

$$i = A_0 + A_1v + A_2v^2 + \dots + A_nv^n, \quad \dots \quad (2)$$

as an approximation to the experimental valve characteristic.

Strutt^{(6), (7)} has approximated to the experimental valve characteristic by a series of the form

$$i = \Sigma A_n e^{b_n v}. \quad \dots \quad (3)$$

For certain special types of characteristic (*e.g.* the anode current—grid voltage characteristic of anode bend detectors), only a very few terms of the series (3) may be required for an accurate analytical approximation. Then, replacing v by the time function for the applied E.M.F., the expression on the right-hand side of equation (3) may be written as a Fourier series, giving the amplitudes and phases of the various harmonics of modulation products in the output current. This writer has considered the applied E.M.F. to consist of one or two sinusoidal components only.

The analytical expression for approximating to the experimentally determined valve characteristic, used in the present paper, has been suggested by Kober⁽⁸⁾ and has the form

$$i = a_0 + av + \sum_{r=1}^{\infty} \left(a_r \cos \frac{2\pi r v}{V} + b_r \sin \frac{2\pi r v}{V} \right). \quad \dots \quad (4)$$

This expression is obtained by making a Fourier analysis of the valve characteristic, less some linear function which approximates to the characteristic, over a range of E.M.F. V , which must be greater than the maximum peak-to-peak value of the applied E.M.F., and must completely cover this range of applied E.M.F. In general, it will be advantageous to choose $a = I/V$, where I is the difference in the values of the currents corresponding to the end points of the range V over which the characteristic is used. The various methods for carrying out the Fourier analysis have been summarized by Carslaw⁽⁹⁾. Using the

analytical form of equation (4) for the valve characteristic, we have here obtained explicit expressions for the harmonic and cross-modulation products in the output current when the applied E.M.F. consists of a single sinusoidal component (Section 2) and when the applied E.M.F. consists of any number of sinusoidal waves of different, but unrelated, frequencies and of arbitrary amplitudes, together with a constant E.M.F. (Section 3).

In Sections 4, 5, and 6 problems involving frequency modulated carrier waves are investigated. Thus, in Section 4, expressions are derived giving the amplitudes and phases of the sideband components of a carrier wave frequency modulated by a number of sinusoidal components of unrelated frequencies. The application of such a wave to a valve of non-linear resistive characteristic, either alone or together with a sinusoidal beating frequency, is made the subject of Sections 5 and 6, and various devices employed in frequency modulation technique, such as amplifiers, frequency multipliers, limiters and frequency changers, are discussed in the light of the results so obtained. In Section 7 we consider the modifications which must be made in the theory if the modulating and carrier frequencies involved are no longer unrelated. It will be apparent that the theory which we have developed for frequency modulated waves is directly applicable to phase modulated waves. A considerable amount of analytical work on frequency modulation has already been published, but most of the writers consider only a single sinusoidal modulating frequency. Exceptions to this are van der Pol ⁽¹⁰⁾, who considers a square modulating wave, and Carson and Fry ⁽¹¹⁾, who have developed from first principles a variable frequency electric circuit theory. It is interesting to note these latter authors consider the analysis of the problem on traditional "steady-state" lines, such as we have used here, to be mathematically intractable.

Although the analysis which has been carried out in this paper is immediately concerned with the application of a wave of E.M.F. to a thermionic valve of non-linear resistive characteristic, it is by no means limited to such use, but may be applied to the solution of many other problems which may be reduced to the same mathematical form.

2. The Application of a Single Sinusoidal Component to a Non-linear Resistive Characteristic.

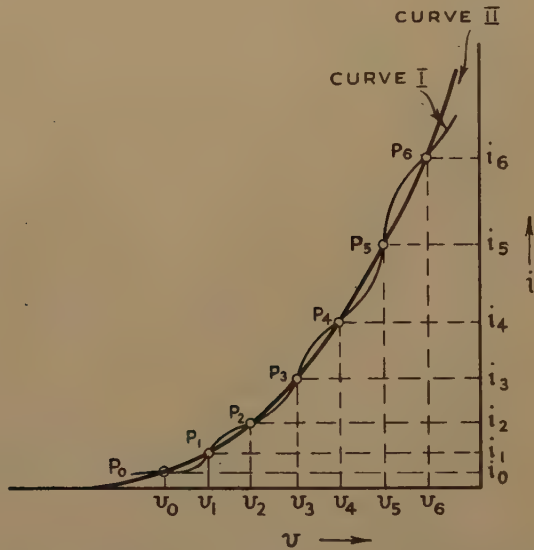
We have seen, in Section 1, that it is customary to approximate to the experimentally determined valve characteristic by an expression of the form (2). The coefficients $A_0, A_1, \dots, A_n$, are usually determined by substituting in equation (2) the corresponding values of i and v for $n+1$ points $P_0, P_1, \dots, P_n$ on the valve characteristic and solving the $n+1$

simultaneous equations which are thus obtained for $A_0, A_1, \dots A_n$. The graphical representation, curve I (fig. 1), of the polynomial so obtained, oscillates, in general, about the experimental characteristic, curve II, cutting it at the points $P_0, P_1, P_2, \dots P_n$. Generally, the accuracy of the approximation increases as the number of points $P_0, P_1, \dots P_n$ increases, but for any given number of points the accuracy of the approximation will depend on the way in which these points are chosen.

Now, suppose we have chosen some polynomial approximation to the valve characteristic of the form (2) and we apply a signal of the form

$$v = V_0 \sin \omega t. \quad \dots \dots \dots (5)$$

Fig. 1.



Approximation to experimental characteristic.

The output current is given by

$$i = \sum_{r=0}^n A_r (V_0 \sin \omega t)^r.$$

This current contains components of angular frequencies $\omega, 2\omega, \dots n\omega$, whose amplitudes are given from this expression, but owing to the fact that the expression is itself only an approximation to the actual valve characteristic, these amplitudes will only approximate to the true amplitudes and we cannot directly determine the accuracy of this approximation.

Now suppose we are concerned with the performance of the valve

over a range V of the characteristic. Then, the valve characteristic may be expressed analytically in the form of equation (4), where

$$a_r = \frac{2}{V} \int_0^V (i - a_0 - av) \cos \frac{2\pi r}{V} \cdot v \, dv$$

$$\text{and} \quad b_r = \frac{2}{V} \int_0^V (i - a_0 - av) \sin \frac{2\pi r}{V} \cdot v \, dv.$$

The output current due to an applied signal of the form (5) may be found by substituting this value for v in the equation (4). Thus,

$$i = a_0 + aV_0 \sin \omega t + \sum_{r=1}^{\infty} \{a_r \cos (rm \sin \omega t) + b_r \sin (rm \sin \omega t)\}, \quad (6)$$

where $m = \frac{2\pi V_0}{V}$.

We can expand each of the expressions $\cos (rm \sin \omega t)$ and $\sin (rm \sin \omega t)$ as a Fourier series. Thus,

$$\cos (rm \sin \omega t)=J_0(rm)+2 \sum_{p=1}^{\infty} J_{2p}(rm) \cos 2p\omega t$$

$$\text{and} \quad \sin (r m \sin \omega t)=2 \sum_{p=1}^{\infty} J_{2 p-1}(r m) \sin (2 p-1) \omega t .$$

Substituting in equation (6), we have

[illegible]

By evaluating the coefficient of the appropriate term in the series on the right-hand side of equation (7) we can find the amplitude of any desired harmonic in the output current. Assuming that the coefficients

$$a_0, a_1, \dots, a_r, \dots, b_1, b_2, \dots, b, \dots$$

have been determined with complete accuracy (which is, of course, not the case in practice), we can obtain an approximation to the amplitude of any given harmonic, which is as accurate as we please, by taking sufficient terms in the particular amplitude we are considering. Thus, if we approximate to the valve characteristic by terminating the series of equation (4) thus,

$$i = a_0 + av + \sum_{r=1}^R \left(a_r \cos \frac{2\pi rv}{V} + b_r \sin \frac{2\pi rv}{V} \right),$$

we see that the calculated coefficients of the terms

$$\cos 2n\omega t \quad \text{and} \quad \sin (2n+1)\omega t$$

in the output current are

$$2 \sum_{r=1}^R a_r J_{2n}(rm) \quad \text{and} \quad 2 \sum_{r=1}^R b_r J_{2n+1}(rm)$$

respectively.

3. *The Application of a Number of Sinusoidal Components of Unrelated Frequencies to a Non-Linear Resistive Characteristic.*

We shall now consider the composition of the output current for an applied E.M.F. of the form given by

$$v = V_0 + \sum_{s=1}^S V_s \cos (\omega_s t + \epsilon_s). \quad \dots \dots \dots (8)$$

Then, substituting this in equation (4), we have

$$\begin{aligned} i = & a_0 + a \left[V_0 + \sum_{s=1}^S V_s \cos \phi_s \right] \\ & + \sum_{r=1}^{\infty} \left[a_r \cos \left\{ r \left(m_0 + \sum_{s=1}^S m_s \cos \phi_s \right) \right\} \right. \\ & \left. + b_r \sin \left\{ r \left(m_0 + \sum_{s=1}^S m_s \cos \phi_s \right) \right\} \right], \quad \dots \dots \dots (9) \end{aligned}$$

where

$$\phi_s = \omega_s t + \epsilon_s \quad \text{for} \quad s = 1, 2, \dots, S$$

and

$$m_s = \frac{2\pi V_s}{V} \quad \text{for} \quad s = 0, 1, 2, \dots, S.$$

Now, the expressions

$$\cos \left\{ r \left(m_0 + \sum_{s=1}^S m_s \cos \phi_s \right) \right\}, \quad \dots \dots \dots (10)$$

and

$$\sin \left\{ r \left(m_0 + \sum_{s=1}^S m_s \cos \phi_s \right) \right\}, \quad \dots \dots \dots (11)$$

are the real and imaginary parts respectively of

$$e^{jr \left(m_0 + \sum_{s=1}^S m_s \cos \phi_s \right)},$$

i. e., of

$$e^{jrm_0} \cdot \prod_{s=1}^S e^{jrm_s \cos \phi_s},$$

i. e., of

$$e^{jrm_0} \cdot \prod_{s=1}^S \left[J_0(rm_s) + 2 \sum_{n=1}^{\infty} j^n J_n(rm_s) \cos n\phi_s \right]. \quad (12)$$

The expression (12) may be expanded as a series whose general term is

$$e^{jrm_0} \cdot 2^N \cdot j^{\alpha} \prod_{s=1}^S [J_{\alpha_s}(rm_s) \cdot \cos \alpha_s \phi_s], \quad (13)$$

where α_s can take any positive integral value or the value zero; $\alpha = \sum_{s=1}^S \alpha_s$ and N is the number of values of s for which α_s is not zero.

Now, the expression $\prod_{s=1}^S (\cos \alpha_s \phi_s)$ may be expanded as the series

$$\frac{1}{2^N} \sum \cos (\pm \alpha_1 \phi_1 \pm \alpha_2 \phi_2 \pm \dots \pm \alpha_S \phi_S),$$

where the summation is carried out over all the permutations of the positive and negative signs. Thus, the expression (12) may be expanded as a series whose general term is

$$e^{jrm_0} \cdot j^{\alpha} \left\{ \prod_{s=1}^S [J_{\alpha_s}(rm_s)] \right\} \cdot \cos \left(\sum_{s=1}^S \alpha_s \phi_s \right), \quad (14)$$

where α_s can now take all positive or negative integral values or the value zero. The expression (14) may be written

$$[j^{\alpha} \cos rm_0 + j^{\alpha+1} \sin rm_0] \left\{ \prod_{s=1}^S [J_{\alpha_s}(rm_s)] \right\} \cos \left(\sum_{s=1}^S \alpha_s \phi_s \right).$$

Thus, the component in the output current of angular frequency

$\left| \sum_{s=1}^S \alpha_s \omega_s \right|$ arising from the term

$$a_r \cos \left\{ r \left(m_0 + \sum_{s=1}^S m_s \cos \phi_s \right) \right\}$$

of expression (9) is

$$2(-1)^{\frac{\alpha}{2}} a_r \cos rm_0 \cdot \left[\prod_{s=1}^S J_{\alpha_s}(rm_s) \right] \cos \left(\sum_{s=1}^S \alpha_s \phi_s \right)$$

if α is even, and

$$2(-1)^{\frac{\alpha+1}{2}} a_r \sin rm_0 \cdot \left[\prod_{s=1}^S J_{\alpha_s}(rm_s) \right] \cos \left(\sum_{s=1}^S \alpha_s \phi_s \right)$$

if α is odd.

These two expressions may be combined in the single expression

$$2a_r \cos\left(rm_0 + \frac{\alpha\pi}{2}\right) \left[\prod_{s=1}^S J_{\alpha_s}(rm_s) \right] \cos\left(\sum_{s=1}^S \alpha_s \phi_s\right),$$

which is valid for α even or odd.

Again, the component in the output current of the same frequency arising from the term

$$b_r \sin\left\{r\left(m_0 + \sum_{s=1}^S m_s \cos \phi_s\right)\right\}$$

of expression (9) is

$$2(-1)^{\frac{\alpha}{2}} b_r \sin rm_0 \left[\prod_{s=1}^S J_{\alpha_s}(rm_s) \right] \cos\left(\sum_{s=1}^S \alpha_s \phi_s\right)$$

if α is even, and

$$2(-1)^{\frac{\alpha-1}{2}} b_r \cos rm_0 \left[\prod_{s=1}^S J_{\alpha_s}(rm_s) \right] \cos\left(\sum_{s=1}^S \alpha_s \phi_s\right)$$

if α is odd.

These two expressions may be combined in the single expression

$$2b_r \sin\left(rm_0 + \frac{\alpha\pi}{2}\right) \left[\prod_{s=1}^S J_{\alpha_s}(rm_s) \right] \cos\left(\sum_{s=1}^S \alpha_s \phi_s\right),$$

which is valid for α even or odd.

Thus, the component of the total output current of angular frequency

$$\left| \sum_{s=1}^S \alpha_s \omega_s \right| \text{ is}$$

$$2 \left[\sum_{r=1}^{\infty} \left\{ a_r \cos\left(rm_0 + \frac{\alpha\pi}{2}\right) + b_r \sin\left(rm_0 + \frac{\alpha\pi}{2}\right) \right\} \left\{ \prod_{s=1}^S J_{\alpha_s}(rm_s) \right\} \right] \cos\left(\sum_{s=1}^S \alpha_s \phi_s\right).$$

This has amplitude

$$2 \left[\sum_{r=1}^{\infty} \left\{ a_r \cos\left(rm_0 + \frac{\alpha\pi}{2}\right) + b_r \sin\left(rm_0 + \frac{\alpha\pi}{2}\right) \right\} \left\{ \prod_{s=1}^S J_{\alpha_s}(rm_s) \right\} \right].$$

Now, defining the phase of any signal $V \cos(\omega t + \epsilon)$ to be ϵ , we see that the phase of the component of the output current of angular frequency

$$\left| \sum_{s=1}^S \alpha_s \omega_s \right| \text{ is}$$

$$\sum_{s=1}^S \alpha_s \epsilon_s \quad \text{or} \quad - \sum_{s=1}^S \alpha_s \epsilon_s$$

according as $\sum_{s=1}^S \alpha_s \omega_s$ is positive or negative.

We see that the amplitude of the component in the output current of angular frequency ω_l —the angular frequency of any single one of the components of the applied signal—is equal to

$$aV_l - 2 \sum_{r=1}^{\infty} \{a_r \sin rm_0 - b_r \cos rm_0\} \left\{ \prod_{s=1}^S J_{\alpha_s}(rm_s) \right\},$$

where $\alpha_s = 1$ for $s = l$ and 0 for all other values of s .

The standing current has the value

$$a_0 + aV_0 + \sum_{r=1}^{\infty} \{a_r \cos rm_0 + b_r \sin rm_0\} \left\{ \prod_{s=1}^s J_0(rm_s) \right\}.$$

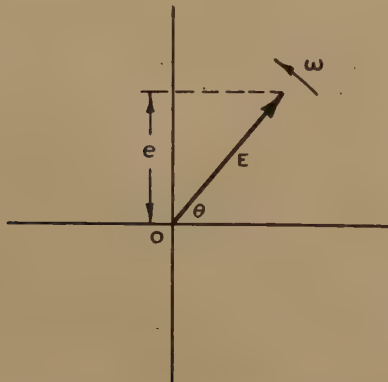
4. The Spectrum of a Frequency Modulated Wave.

A similar mathematical method to that used in Section 3 may be applied to the calculation of the components of the spectrum produced when a carrier wave is frequency modulated by a signal consisting of a number of sinusoidal components having unrelated frequencies.

Suppose the instantaneous angular frequency of the modulated carrier wave is given by

$$\omega = \omega_0 \left[1 + \sum_{s=1}^s k_s \sin(\omega_s t + \epsilon_s) \right], \quad \dots \dots \dots (15)$$

Fig. 2.



Vectorial representation of frequency modulated wave of E.M.F.

where ω_0 is the angular frequency of the unmodulated carrier wave, and $\omega_1, \omega_2, \dots, \omega_s$ are the angular frequencies of the components of the modulating wave.

The instantaneous angle of displacement θ of the carrier vector is given by

$$\theta = \int \omega \cdot dt. \quad \dots \dots \dots (16)$$

The modulated carrier E.M.F. may be written

$$e = E \sin \theta, \quad \dots \dots \dots (17)$$

as can be seen from fig. 2, where E is the peak amplitude of E.M.F. and e is the instantaneous value.

Now, combining equations (15), (16) and (17), we have

$$\begin{aligned} e &= E \sin \left[\omega_0 t - \sum_{s=1}^S \frac{k_s}{\omega_s} \cdot \omega_0 \cos (\omega_s t + \epsilon_s) \right] \\ &= E \sin \left[\omega_0 t + \sum_{s=1}^S m_s \cos \phi_s \right], \quad \dots \quad (18) \end{aligned}$$

where $\phi_s = \omega_s t + \epsilon_s$ and $m_s = -\frac{k_s}{\omega_s} \cdot \omega_0$.

The expression (18) is identical in form with expression (11), where $r=1$ and $m_0 = \omega_0 t$ and is thus, from expression (14), equal to the imaginary part of a series whose general term is

$$E e^{j\omega_0 t} \cdot j^\alpha \left\{ \prod_{s=1}^S J_{\alpha_s}(m_s) \right\} \cos \left(\sum_{s=1}^S \alpha_s \phi_s \right),$$

where α_s can take all positive or negative integral values, or the value zero.

We thus see that the expression (18) may be expanded as a series whose general term is

$$E \cdot \left\{ \prod_{s=1}^S J_{\alpha_s}(m_s) \right\} \cos \left(\omega_0 t + \sum_{s=1}^S \alpha_s \phi_s + \overline{\alpha-1} \cdot \frac{\pi}{2} \right) \quad \dots \quad (19)$$

The amplitude of the sideband of angular frequency $\left(\omega_0 + \sum_{s=1}^S \alpha_s \omega_s \right)$ is

$$E \left\{ \prod_{s=1}^S J_{\alpha_s}(m_s) \right\}, \quad \dots \quad (20)$$

and its phase is

$$\sum_{s=1}^S \alpha_s \epsilon_s + \overline{\alpha-1} \cdot \frac{\pi}{2}.$$

In cases of practical interest we shall not generally be concerned with sideband components arising from terms for which $\left(\omega_0 + \sum_{s=1}^S \alpha_s \omega_s \right)$ is negative, but it will be apparent how to deal with such cases from Section 3.

The amplitudes of the sidebands produced, when a carrier wave is modulated in frequency by two sinusoidal terms at the same time, have been evaluated by Crosby ⁽¹²⁾, and the above general solution may be seen to be in agreement with his special case.

From expression (20), it is seen that as the number of modulating terms is increased the amplitudes of the sidebands are decreased ($J_\alpha(m)$ is always fractional), as may be expected, since the total energy of the modulated wave is independent of the form of the modulating wave of E.M.F.

Since the sidebands extend from zero to infinite frequency, we cannot

give any definite band width for the frequency modulated wave, but, in general, beyond a certain frequency range from the carrier, the sideband amplitudes are negligibly small and so have an inappreciable effect on the resulting frequency modulated wave.

Thus, it is well known that in the case of the frequency modulation of a carrier wave by a single modulating term of angular frequency ω_s , the resulting spectrum of sideband components is concentrated into an angular frequency band, the width of which, on either side of the carrier components, is slightly larger than the angular frequency modulation excursion $-k_s\omega_0$, provided m_s is large compared with unity. The amplitudes of the sidebands, whose frequencies lie outside this range, rapidly approach zero. This depends on the fact that $J_{\alpha_s}(m_s)$ approaches zero rapidly, as α_s increases beyond m_s , if $m_s \gg 1$.

When we have a number of simultaneously applied modulating components, we see from the formula (20) that the amplitudes become very small for components whose angular frequencies lie much outside the band $\omega_0 \pm \sum_{s=1}^S m_s \omega_s$, at any rate if each value of m_s is fairly large compared with unity. Then, the half-band width is given approximately by $\omega_0 \sum_{s=1}^S (-k_s)$.

5. The Application of a Frequency Modulated Wave to a Non-Linear Resistive Characteristic.

We shall here find the effect of the non-linearity of the characteristic on the harmonic content of the output current when the input signal is the frequency modulated wave of E.M.F. given in equation (18). Suppose the equation of the characteristic is given by equation (4). Then, substituting $v=e$ in the equation and putting

$$\omega_0 t + \sum_{s=1}^S m_s \cos \phi_s = \theta \quad \text{and} \quad \frac{2\pi E}{V} = m,$$

we see, from equation (7), that the output current is given by a series

$$Q_0 + R_0 \sin \theta + \sum_{p=1}^{\infty} \{Q_p \cos 2p\theta + R_p \sin 2p+1 \cdot \theta\}, \quad \dots \quad (21)$$

where

$$Q_0 = a_0 + \sum_{r=1}^{\infty} a_r J_0(rm),$$

$$R_0 = aE + 2 \sum_{r=1}^{\infty} b_r J_1(rm),$$

$$Q_p = 2 \sum_{r=1}^{\infty} a_r J_{2p}(rm)$$

and

$$R_p = 2 \sum_{r=1}^{\infty} b_r J_{2p+1}(rm).$$

We can see, by an argument similar to that of Section 4, that $\cos 2p\theta$ is equal to the real part of a series whose general term is

$$e^{j \cdot 2p\omega_0 t} \cdot j^\alpha \left\{ \prod_{s=1}^S J_{\alpha_s}(2pm_s) \right\} \cos \left(\sum_{s=1}^S \alpha_s \phi_s \right), \quad \dots \quad (22)$$

and that $\sin (2p+1)\theta$ is equal to the imaginary part of a series whose general term is

$$e^{j \cdot (2p+1)\omega_0 t} \cdot j^\alpha \left\{ \prod_{s=1}^S J_{\alpha_s}(\overline{2p+1} \cdot m_s) \right\} \cos \left(\sum_{s=1}^S \alpha_s \phi_s \right). \quad \dots \quad (23)$$

It appears, from the expression (22) and (23), that the series (21) may be expanded in terms of sidebands of angular frequencies

$$\left(2p\omega_0 + \sum_{s=1}^S \alpha_s \omega_s \right) \quad \text{and} \quad \left(\overline{2p+1} \cdot \omega_0 + \sum_{s=1}^S \alpha_s \omega_s \right),$$

and that the former sideband can arise only from the term $Q_p \cos 2p\theta$, and the latter from the term $R_p \sin (2p+1)\theta$.

From this, together with the expression (22), we see that the sideband of angular frequency $\left(2p\omega_0 + \sum_{s=1}^S \alpha_s \omega_s \right)$ has amplitude

$$Q_p \left\{ \prod_{s=1}^S J_{\alpha_s}(2pm_s) \right\} \quad \text{and phase} \quad \sum_{s=1}^S \alpha_s \epsilon_s + \alpha \cdot \frac{\pi}{2}.$$

Again, applying the same considerations to expressions (23), we see that the sideband of angular frequency $\left(\overline{2p+1} \cdot \omega_0 + \sum_{s=1}^S \alpha_s \omega_s \right)$ has amplitude

$$R_p \left\{ \prod_{s=1}^S J_{\alpha_s}(\overline{2p+1} \cdot m_s) \right\} \quad \text{and phase} \quad \sum_{s=1}^S \alpha_s \epsilon_s + (\alpha-1) \frac{\pi}{2}.$$

It can be seen from equation (21) that the output current may be considered to consist of a number of waves of current of the form

$$\frac{\cos}{\sin} \left[r \left\{ \omega_0 t + \sum_{s=1}^S m_s \cos (\omega_s t + \epsilon_s) \right\} \right],$$

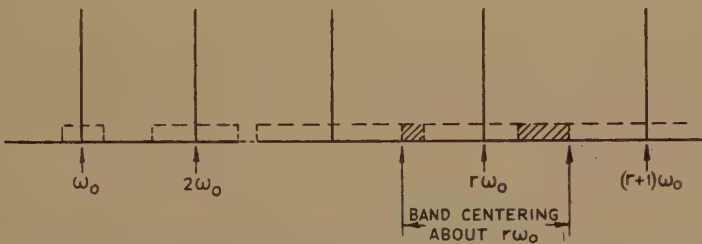
where the cosine is taken when r is even (i. e., $r=2p$) and the sine is taken when r is odd (i. e., $r=2p+1$). Now, each of these waves is a wave which could be obtained by frequency modulating a carrier of angular frequency $r\omega_0$ with the same modulating signal, the depth of modulation for each component being proportional to r . If $rm_s \gg 1$, for all values of s from 1 to S ,

$$\frac{\cos}{\sin} \left[r \left\{ \omega_0 t + \sum_{s=1}^S m_s \cos (\omega_s t + \epsilon_s) \right\} \right]$$

has sideband components of appreciable amplitude lying in, or just outside, the band of angular frequencies $r\omega_0 \pm \sum_{s=1}^S rm_s \omega_s$ (from Section 4)

If the condition $rm_s \gg 1$ is not fulfilled for all values of s from 1 to S , then the width of the band in which the sidebands of appreciable amplitude lie is generally larger than that given by the angular frequency limits $r\omega_0 \pm \sum_{s=1}^S rm_s \omega_s$. We see then that no matter how small are the initial depths of modulation, by taking r sufficiently large we reach a point where the band-width formed about the r th harmonic of the carrier overlaps with that formed about the $(r+1)$ th harmonic, as shown in fig. 3. However, if for some value of r this does not occur, we can by the use of suitable selective circuits filter out the frequency modulated component whose carrier wave has angular frequency $r\omega_0$. This wave will carry the modulating signal in a substantially undistorted form. Again, it should be noted that even if the frequency band formed about the r th harmonic of the carrier overlaps with those formed about the $(r+1)$ th and $(r-1)$ th harmonics, it may still be possible to carry the

Fig. 3.



Overlapping of harmonic signals in a frequency multiplier stage.

frequency multiplication as far as the angular frequency $r\omega_0$, provided the non-linear characteristic has the correct form, viz., if $r=2p$, R_{p-1} and R_p are small compared with Q_p , or if $r=2p+1$, Q_p and Q_{p+1} are small compared with R_p .

We see also that it will be possible, subject to certain limitations of the modulating frequencies and depths of modulation, to filter from the output a substantially undistorted reproduction of the input frequency modulated E.M.F.

Advantage is taken of the fact that a frequency modulated carrier wave may be multiplied in frequency a number of times, without introducing distortion of the modulation form, in frequency modulation communication technique. Some experiments have been described by Armstrong⁽¹³⁾, who, although suggesting that a multiplying factor of 500–800 times is more usual, has obtained factors as high as 10,000. It is clear that, owing partly to the overlapping of the signals centering about the r th and the $(r+1)$ th, harmonics of the carrier frequency,

at some value of r , and partly to the decreasing amplitudes of the harmonics when r increases, there is a limit in the multiplication factor of a single valve. Thus in practice a number of multiplying stages in cascade may be used.

6. Frequency Changing a Frequency Modulated Wave.

Suppose now that the frequency modulated carrier wave of E.M.F. of equation (18) is applied to a valve having a characteristic given by equation (4), together with a sinusoidal E.M.F. $E' \cos(\omega' t + \epsilon')$, ($= E' \cos \phi'$) and a steady potential E_0 , where the angular frequencies ω' , ω_0 , and ω_1 , $\omega_2, \dots, \omega_s$, are unrelated.

Then, the total E.M.F. applied to the valve may be written

$$v = E_0 + E' \cos \phi' + E \cos \phi,$$

where

$$\phi = \omega_0 t + \sum_{s=1}^s m_s \cos \phi_s - \frac{\pi}{2}.$$

Substituting this expression in equation (4), and following the argument of Section (3), we see that the output current may be expressed as a series whose general term (apart from the constant and linear terms) is

$$\sum_{r=1}^{\infty} \left[\left\{ a_r \cos \left(rm_0 + \overline{\beta + \beta'} \cdot \frac{\pi}{2} \right) + b_r \sin \left(rm_0 + \overline{\beta + \beta'} \cdot \frac{\pi}{2} \right) \right\} \{ J_\beta(rm) J_{\beta'}(rm') \} \right] \cos(\beta\phi + \beta'\phi'), \quad (24)$$

where β and β' have any integral values or the value zero.

$$\text{Also,} \quad \frac{2\pi E_0}{V} = m_0, \quad \frac{2\pi E'}{V} = m', \quad \text{and} \quad \frac{2\pi E}{V} = m.$$

Now, the term $\cos(\beta\phi + \beta'\phi')$ which is equal to $\sin\left(\beta\phi + \beta'\phi' + \frac{\pi}{2}\right)$ can be expanded, from Section 4, in a series whose general term is

$$\left\{ \prod_{s=1}^s J_{\alpha_s}(\beta m_s) \right\} \cos \left(\beta' \phi' - \overline{\beta - 1} \frac{\pi}{2} + \beta \omega_0 t + \sum_{s=1}^s \alpha_s \phi_s + \overline{\alpha - 1} \frac{\pi}{2} \right).$$

We thus see that the component in the output current of angular frequency $\left| \beta \omega_0 + \beta' \omega' + \sum_{s=1}^s \alpha_s \omega_s \right|$ has amplitude

$$2 \sum_{r=1}^{\infty} \left[\left\{ a_r \cos \left(rm_0 + \overline{\beta + \beta'} \frac{\pi}{2} \right) + b_r \sin \left(rm_0 + \overline{\beta + \beta'} \frac{\pi}{2} \right) \right\} \{ J_\beta(rm) \cdot J_{\beta'}(rm') \} \right] \left\{ \prod_{s=1}^s J_{\alpha_s}(\beta m_s) \right\}$$

and phase $\left(\beta' \epsilon' + \overline{\alpha - \beta} \frac{\pi}{2} + \sum_{s=1}^s \alpha_s \epsilon_s \right).$

From expression (24), it appears that the output current may be considered to consist of a number of carrier waves of angular frequency $|\beta\omega + \beta'\omega'|$, where β and β' are positive or negative integers or zero, each carrier being modulated with the same signal, the depth of modulation depending on the frequency of the carrier. In general, the frequency band occupied by one of these modulated carriers will be overlapped by the frequency bands of other modulated carriers, owing to the fact that $(\beta\omega_0 + \beta'\omega')$ may have nearly equal values for two or more pairs of values of β, β' .

However, if the amplitude in the expression of the form (24) for the overlapping band is small, then the amount of distortion introduced is negligible. This occurs if β or β' is greater than rm or rm' respectively, for small values of r , where β and β' are the numbers of the harmonics of the angular frequencies ω_0, ω' in the carrier with which the interfering frequency modulated wave is formed. It should be noted that, in general, the coefficients a_r, b_r will be small for large values of r , making the contribution of these terms to the sideband amplitudes small.

7. Extension to the Case when the Frequencies are Related.

In Section 6 we have limited our discussion to the case when the angular frequencies $\omega_0, \omega', \omega_1, \omega_2, \dots, \omega_s$ are unrelated, *i. e.*, there exists no linear relation between them of the form

$$x_0\omega_0 + x'\omega' + \sum_{s=1}^S x_s\omega_s = 0, \quad . \quad . \quad . \quad . \quad . \quad (25)$$

where all the x 's are positive or negative integers or zero.

If this is not true then relations of the form (25) may exist between the angular frequencies. However, it is clear that not more than $(S+1)$ independent relations of this form can exist between these $(S+2)$ angular frequencies.

Suppose that, in fact, L such independent relations exist. We shall denote them by

$$x_{0l}\omega_0 + x_l'\omega' + \sum_{s=1}^S x_{sl}\omega_s = 0, \quad . \quad . \quad . \quad . \quad . \quad (26)$$

where $l=1, 2, \dots, L$, and the x 's may be positive or negative integers or zero, and have no common integral factor for any given value of l .

Then, the angular frequency

$$\begin{aligned} \beta_0\omega_0 + \beta'\omega' + \sum_{s=1}^S \alpha_s\omega_s = & \left(\beta_0\omega_0 + \beta'\omega' + \sum_{s=1}^S \alpha_s\omega_s \right) \\ & + \sum_{l=1}^L K_l \left(x_{0l}\omega_0 + x_l'\omega' + \sum_{s=1}^S x_{sl}\omega_s \right), \end{aligned}$$

where $K_1, K_2, \dots, K_L$ can take any set of positive or negative integral values or the value zero.

Thus, the angular frequency

$$\begin{aligned}
 \beta_0 \omega_0 + \beta' \omega' + \sum_{s=1}^S \alpha_s \omega_s \\
 &= \left(\beta_0 + \sum_{l=1}^L K_l x_{0l} \right) \omega_0 + \left(\beta' + \sum_{l=1}^L K_l x_{l'} \right) \omega' \\
 &+ \left(\alpha_1 + \sum_{l=1}^L K_l x_{1l} \right) \omega_1 + \left(\alpha_2 + \sum_{l=1}^L K_l x_{2l} \right) \omega_2 \\
 &+ \dots \quad \quad \quad + \dots \\
 &+ \dots \quad \quad \quad + \left(\alpha_S + \sum_{l=1}^L K_l x_{Sl} \right) \omega_S.
 \end{aligned}$$

Each set of values of $K_1, K_2, \dots, K_L$ gives rise to a component in the output current of angular frequency

$$\left(\beta_0 \omega_0 + \beta' \omega' + \sum_{s=1}^S \alpha_s \omega_s \right),$$

whose amplitude and phase can be written down from Section 6. We shall denote this amplitude and phase by $A_{K_1 K_2 \dots K_L}$ and $\epsilon_{K_1 K_2 \dots K_L}$ respectively.

The amplitude of resultant component in the output current of angular frequency $\left(\beta_0 \omega_0 + \beta' \omega' + \sum_{s=1}^S \alpha_s \omega_s \right)$ may thus be written as

$$\{ (A_{K_1 K_2 \dots K_L} \cos \epsilon_{K_1 K_2 \dots K_L})^2 + (A_{K_1 K_2 \dots K_L} \sin \epsilon_{K_1 K_2 \dots K_L})^2 \}^{\frac{1}{2}},$$

and the phase as

$$\tan^{-1} \frac{A_{K_1 K_2 \dots K_L} \sin \epsilon_{K_1 K_2 \dots K_L}}{A_{K_1 K_2 \dots K_L} \cos \epsilon_{K_1 K_2 \dots K_L}},$$

where the double suffix notation for summation is used and the summations are carried out for all values of $K_1, K_2, \dots, K_L$ from $-\infty$ to $+\infty$.

In practice, we need only consider sets of values of $K_1, K_2, \dots, K_L$, for which $\left(\alpha_s + \sum_{l=1}^L K_l x_{sl} \right)$ is not much greater than βm_s for all values of s from 1 to S inclusive.

Although the discussion in this section has been restricted to the problem of the frequency changing of a frequency modulated carrier wave, it will be apparent how the problems of Sections 3, 4, and 5 may be dealt with when the frequencies involved are no longer unrelated.

We see that the general conclusions reached in Section 4 regarding the band-width and in Sections 5 and 6 regarding the distortion introduced by amplifying and frequency changing a frequency modulated wave still apply when the frequencies involved are mutually related.

References.

- (1) A. C. Bartlett, "The Calculation of Modulation Products." *Phil. Mag.* xvi. p. 845 (Oct. 1933).
- (2) A. C. Bartlett, "The Calculation of Modulation Products.—II." *Phil. Mag.* xvii. p. 628 (March 1934).
- (3) D. C. Espley, "The Calculation of Harmonic Production in Thermionic Valves with Resistive Loads." *Proc. I. R. E.* xxi. no. 10 (October, 1933).
- (4) D. C. Espley, "Harmonic Production and Cross Modulation in Thermionic Valves with Resistive Loads." *Proc. I. R. E.* xxii. no. 6 (June, 1934).
- (5) D. C. Espley, "Harmonic Analysis by the Method of Central Differences." *Phil. Mag. ser. 7*, xxviii. p. 338 (Sept. 1939).
- (6) M. J. O. Strutt, "Anode Bend Detection." *Proc. I. R. E.* xxiii. p. 945 (Aug. 1935).
- (7) M. J. O. Strutt, "On Conversion Detectors." *Proc. I. R. E.* xxii. p. 981 (Aug. 1934).
- (8) C. L. Kober, "Die Berechnung von Nichtlinearen Verzerrungen." *Elek. Nach. Tech.* xiii. p. 336 (1936).
- (9) Carslaw, 'Fourier's Series and Integrals.' Appendix I. Macmillan & Co. (1930).
- (10) Van der Pol, "Frequency Modulation." *Proc. I. R. E.* xviii. p. 1194 (July 1930).
- (11) Carson and Fry, "Variable Frequency Electric Circuit Theory, with Application to the Theory of Frequency Modulation." *B. S. T. J.* xvi. p. 513 (1937).
- (12) M. G. Crosby, "Carrier and Side Frequency Relations, with Multitone Frequency or Phase Modulation." *R. C. A. Review*, cxi. (July, 1938).
- (13) E. H. Armstrong, "Method of Reducing Disturbance in a Radio Signal by a System of Frequency Modulation." *Proc. I. R. E.* xxiv. p. 689 (1936).

XXIII. *Thermal Stresses in Long Cylindrical Bodies.*

By B. E. GATEWOOD †.

[Received July 26, 1940.]

ABSTRACT.

The problem of finding thermal stresses and displacements in long cylinders is reduced to obtaining a particular integral for the Poisson equation $\nabla^2 V = kT$, where T is the temperature distribution and k is a constant, and to solving the biharmonic equation $\nabla^4 U = 0$ when the first derivatives of the function U are known on the boundary of the cross-section of the body. These two differential equations are related by the conditions that on the boundary the first derivatives of U are given in terms of the first derivatives of V . The biharmonic equation is solved by use of two analytic functions [Muschelisvili, 'Bulletin de l'Academie des Sciences de Russie, IV Serie, Tome 13, 2nd partie, 1919; 'Mathematische Annalen,' vol. cvii. 1933]. The region within which the solution is desired is mapped conformally on the unit circle by a rational function, and from the boundary values on the unit circle the required analytic functions are determined directly by the use of the Cauchy integral formula. The methods are extended to multi-connected bodies and to bodies composed of several different materials. Special problems of the circular cylinder, the hollow circular cylinder, the concentric and eccentric composite circular cylinders are solved.

1. *Introduction.*

THE problem is to find the stresses and displacements in a long cylindrical body subjected to a change of temperature which depends on the co-ordinates of the cross-section of the body but is independent of the co-ordinate of the length of the body. In the literature, throughout the last one hundred years, there have appeared about twenty papers on this problem. Duhamel ⁽¹⁾, Alibandi ⁽²⁾, Leon ⁽³⁾, Lorenz ⁽⁴⁾, A. Foppl ⁽⁵⁾, Lees ⁽⁶⁾, Dahl ⁽⁷⁾, Barker ⁽⁸⁾, Timoshenko ⁽⁹⁾, and Den Hartog ⁽¹⁰⁾ gave discussions of the long circular cylinder and the long hollow circular cylinder, with the temperature a function of the radius and of the time. Eichelberg ⁽¹¹⁾ considered the hollow cylinder for the case of periodic

† Communicated by the Author.

temperature variation. Poritsky ⁽¹²⁾ examined the circular cylinder of two materials concentric to each other with the temperature change a constant, and also the hollow circular cylinder under the hypothesis of a steady state heat distribution. Lebedew ⁽¹³⁾ solved the problem of the circular cylinder for the case of the temperature a function of the radius and of the angle about the centre. Goodier ⁽¹⁴⁾ and Biot ⁽¹⁵⁾ considered theoretical aspects of the problem as well as some cases of the hollow circular cylinder. From these remarks it is seen that the circular cylinder and the hollow circular cylinder are the only bodies for which explicit expressions for the stresses and displacements have been given; that the temperature distribution has been used as a function of the radius in most cases; and that there is only one paper on the composite body. Further, it should be noted that in most of these papers the authors have regarded the problem as an ordinary stress problem with given body and surface forces arising from the effects due to the temperature change in the body.

In this paper the equilibrium conditions and boundary conditions of elasticity theory will be used as they stand without making modifications to exhibit the temperature effects as body and surface forces. The purpose of this paper is to develop a theory for the problem of thermal stresses in long cylindrical bodies which will include all the cases mentioned as well as cases which are more general with respect to temperature distribution, body cross-section, body connectivity, and body composition. Furthermore, the purpose is to provide a method of solution which will yield explicit results for the stresses and displacements. To accomplish this latter purpose, restrictions which limit the mathematical theory are imposed from time to time. Applications of the method are made to several special problems already solved by different methods to indicate its brevity and directness, as well as to two particular problems not in the literature to the author's knowledge.

2. Derivation of the Differential Equation.

Consider a cylindrical body whose length is large compared to the greatest linear dimension of its cross-section. Take the length of the body in the direction of the z -axis and let the cross-section be the xy -plane. The body is assumed to be subjected to a change of temperature which is a function of the co-ordinates x and y , but which is independent of z . This change of temperature will cause the body to undergo an expansion which generally cannot proceed freely, so that stresses are set up. Let the stresses in the z -direction be balanced by suitably distributed forces over the ends of the body so that the cross-sections remain plane. These distributed forces will give a resultant force and a couple acting on the

body; the problem of compression due to the resultant force and the problem of bending due to the moment of the couple can be solved separately and superposed upon the solution obtained by using plane sections. By Saint Venant's Principle ⁽¹⁶⁾ the results for the stresses and displacements obtained by this superposition of solutions will hold at a sufficient distance from the ends of the body. In the following discussion the strain in the z -direction will be adjusted as a constant of such value that the resultant normal force over the cross-section is zero, whence the problem of compression need not be brought in. Stress distributions near the ends of the cylinder will not be considered in this paper.

Let $T_0(x, y) = T(x, y, t_0)$ be the initial temperature of the body, and let $T(x, y, t)$ be the temperature at a later time diminished by T_0 . Then the Duhamel-Neumann law connecting stress, strain, and displacements is as follows †:

$$e_{ij} = \frac{1}{2} \left(\frac{\partial u_j}{\partial x_i} + \frac{\partial u_i}{\partial x_j} \right) = \frac{1}{2G} \left[2\sigma_{ij} + \left(2G\alpha T - \sigma_{ii} - \frac{s}{m+1} \right) \delta_{ij} \right], \quad (2.1)$$

where $i, j = 1, 2, 3$, δ_{ij} is the Kronecker delta, $s = \sum_{i=1}^3 \sigma_{ii}$, $G = \frac{Em}{2(m+1)}$, E is Young's modulus, m is the reciprocal of Poisson's ratio, and α is the coefficient of linear thermal expansion. Throughout the discussion E , m , and α shall be regarded as constant. Under conditions on the body given above it follows that $e_{23} \equiv \gamma_{yz} = 0$, $e_{31} \equiv \gamma_{zx} = 0$, $e_{33} = e_z = \text{constant}$. Assume all body forces to be zero, so that the equilibrium conditions are as follows:

$$\sum_{j=1}^2 \frac{\partial \sigma_{ij}}{\partial x_j} = 0, \quad i = 1, 2. \quad (2.2)$$

The conditions to be fulfilled on the surface of the body are

$$\sum_{j=1}^2 \sigma_{ij} \cos(n x_j) = X_i, \quad i = 1, 2, \quad (2.3)$$

where X_1 and X_2 are the components of applied surface forces on the lateral surface of the body in the direction of the x and y axes, respectively; and where n is the exterior normal to the lateral surface. From the assumption made above that the constant strain e_z is to be determined by taking the resultant force in the z -direction to be zero, it follows that for the cross-section A of the body there arises the condition

$$\int_A \sigma_z dx dy = 0. \quad (2.4)$$

† For notation see Trefftz, 'Handbuch der Physik,' vol. vi. chapter 2. Here 1, 2, 3 correspond to the x, y, z co-ordinates respectively, $x_1, x_2, x_3 = x, y, z$, $u_1, u_2, u_3 = u, v, w$, etc.

The compatibility condition is ⁽¹⁶⁾

$$\frac{\partial^2 e_x}{\partial y^2} + \frac{\partial^2 e_y}{\partial x^2} = \frac{\partial^2 \gamma_{xy}}{\partial x \partial y} \quad (2.5)$$

The equations (2.1), (2.2), (2.3), (2.4), and (2.5) form a system of equations for determining the displacements, strains, and stresses of the body under consideration. In the group of equations (2.1) solve the equation in which $i=j=3$ for $\sigma_z = \sigma_{33}$ and substitute the result in the other equations containing σ_z . This divides the group (2.1) into two parts :

$$\left. \begin{aligned} e_{ij} = \frac{1}{2} \left(\frac{\partial u_j}{\partial x_i} + \frac{\partial u_i}{\partial x_j} \right) &= \frac{1}{2G} \left[2\sigma_{ij} - \left(\sigma_{ij} + \frac{s - \sigma_z}{m} \right) \delta_{ij} \right] \\ &- \left(\frac{e_z}{m} - \frac{m+1}{m} \alpha T \right) \delta_{ij}, \quad i, j = 1, 2, \end{aligned} \right\} \quad (2.6)$$

$$\left. \begin{aligned} \sigma_z &= E(e_z - \alpha T) + \frac{1}{m}(\sigma_x + \sigma_y), \\ e_z &= \text{constant}, \quad \gamma_{yz} = \frac{1}{G} \tau_{yz} = 0, \quad \gamma_{zx} = \frac{1}{G} \tau_{zx} = 0. \end{aligned} \right\} \quad (2.7)$$

The discussion hereafter will be devoted to the solution of the two-dimensional system of equations (2.2), (2.3), (2.5), and (2.6), since when this system is solved the system consisting of equations (2.4) and (2.7) can be solved.

It is well known that a stress function $\mathbf{O}$ defined by

$$\sigma_x = \frac{\partial^2 \mathbf{O}}{\partial y^2}, \quad \sigma_y = \frac{\partial^2 \mathbf{O}}{\partial x^2}, \quad \tau_{xy} = - \frac{\partial^2 \mathbf{O}}{\partial x \partial y} \quad (2.8)$$

satisfies the equilibrium conditions (2.2). Substitution of this stress function into equations (2.6) and substitution of the result into equation (2.5) gives the following fourth order differential equation for the stress function $\mathbf{O}$:

$$\nabla^4 \mathbf{O} + \frac{Em\alpha}{m-1} \nabla^2 T = 0. \quad (2.9)$$

Let $\mathbf{O} = \mathbf{U} - \mathbf{V}$, where $\mathbf{V}$ is any solution of the differential equation $\nabla^2 \mathbf{V} = kT$; then the differential equation (2.9) is equivalent to a system of two differential equations

$$\nabla^2 \mathbf{V} = kT, \quad \nabla^4 \mathbf{U} = 0, \quad \text{where } k = \frac{Em\alpha}{m-1}. \quad (2.10)$$

In terms of the functions U and V the surface conditions (2.3) become

$$\left. \begin{aligned} \frac{\partial^2 U}{\partial x_i \partial x_i} \cos (n x_j) - \frac{\partial^2 U}{\partial x_i \partial x_j} \cos (n x_i) &= X_j + \frac{\partial^2 V}{\partial x_i \partial x_i} \cos (n x_j) \\ - \frac{\partial^2 V}{\partial x_i \partial x_j} \cos (n x_i), \quad i, j &= 1, 2, i \neq j. \end{aligned} \right\} \quad (2.11)$$

Thus the problem has been reduced to solving the equations (2.10) subject to the conditions (2.11). The function V will be called the heat function, while the function U will be called the biharmonic function. It will be necessary to find the heat function V first in order to obtain the boundary conditions (2.11) for the biharmonic function U . It should be observed that only a particular solution of $\nabla^2 V = kT$ is necessary to give the function V , since any solution of $\nabla^2 V = 0$ can be taken care of by a harmonic function $\dagger$ arising in the solution of the biharmonic equation for U .

In terms of the functions U and V the stresses, by (2.8), are

$$\sigma_x = \frac{\partial^2 U}{\partial y^2} - \frac{\partial^2 V}{\partial y^2}, \quad \sigma_y = \frac{\partial^2 U}{\partial x^2} - \frac{\partial^2 V}{\partial x^2}, \quad \tau_{xy} = \frac{\partial^2 V}{\partial x \partial y} - \frac{\partial^2 U}{\partial x \partial y}. \quad (2.12)$$

From (2.12) and (2.6) the equations for the displacements are ⁽¹⁷⁾:

$$u_i = \frac{2(m-1)}{mG} p_i - \frac{1}{2G} \left(\frac{\partial U}{\partial x_i} - \frac{\partial V}{\partial x_i} \right) - \frac{e_z x_i}{m}, \quad i=1, 2, \quad \dots \quad (2.13)$$

where

$$p_1 + i p_2 = \mathbf{O}_1(z) = \frac{1}{4} \int_c^z f(z) dz. \quad \dots \quad (2.14)$$

Here $f(z) = \nabla^2 U + iQ(x, y)$, where $Q(x, y)$ is the function conjugate to the harmonic function $\nabla^2 U$; $f(z)$ and $\mathbf{O}_1(z)$ are analytic, and c is the initial point on the path of integration in the region which is assumed simply connected.

3. The Biharmonic Function U .

Assume that a particular integral of the Poisson equation $\nabla^2 V = kT$ has been obtained so that the heat function V is known. Then the function U can be obtained by solving the differential equation $\nabla^4 U = 0$, subject to the conditions (2.11) which can be written in the form

$$\frac{d}{ds} \frac{\partial U}{\partial x} = -Y + \frac{d}{ds} \frac{\partial V}{\partial x}, \quad \frac{d}{ds} \frac{\partial U}{\partial y} = X + \frac{d}{ds} \frac{\partial V}{\partial y}, \quad \dots \quad (3.1)$$

or, if these derivatives are single-valued, in the form

$$\frac{\partial U}{\partial x} + i \frac{\partial U}{\partial y} = i \int_{s_0}^s (X + iY) ds + \frac{\partial V}{\partial x} + i \frac{\partial V}{\partial y}, \quad \dots \quad (3.2)$$

$\dagger$ This remark will be clarified when the general solution of the biharmonic equation is considered in section 3.

where s is the arc-length on the boundary C of the cross-section S of the body, with C a simple curve enclosing the simply-connected region S . The problem of solving the biharmonic equation when the first derivatives are known on the boundary has received a great deal of study. In particular, Muschelisvili has developed several methods of solution, and it is the first method developed by Muschelisvili ^{(17), (18)} that will be used in the following discussion; this method has certain restrictions that do not arise in the subsequent one ⁽¹⁹⁾, but it is easy to use in making calculations and in obtaining explicit results.

In terms of the analytic functions $\mathbf{O}_1(z)$ and $\mathbf{U}_2(z)$ the function U can be written as

$$2U = \bar{z}\mathbf{O}_1(z) + z\bar{\mathbf{O}}_1(\bar{z}) + \mathbf{U}_2(z) + \bar{\mathbf{U}}_2(\bar{z}), \quad (3.3)$$

where the function $\mathbf{O}_1(z)$ is the same as the function defined in equation (2.14). Define $\mathbf{U}_1(z) = \mathbf{U}_2'(z)$ and differentiate (3.3):

$$\frac{\partial U}{\partial x} + i \frac{\partial U}{\partial y} = \mathbf{O}_1(z) + z\bar{\mathbf{O}}_1'(\bar{z}) + \bar{\mathbf{U}}_1(\bar{z}), \quad (3.4)$$

which is valid in the region S . Now suppose that the region S , bounded by a simple closed curve C drawn in the z -plane, admits a conformal representation on a circle S' of radius unity and circumference L in the complex s -plane by means of a rational function $z=w(s)$ †. Substitution of $z=w(s)$ in (3.4) gives

$$\frac{\partial U}{\partial x} + i \frac{\partial U}{\partial y} = \mathbf{O}(s) + \frac{w(s)}{w'(\bar{s})} \bar{\mathbf{O}}'(\bar{s}) + \bar{\mathbf{U}}(\bar{s}), \quad (3.5)$$

where $\mathbf{U}(s) = \mathbf{U}_1[w(s)]$, $\mathbf{O}(s) = \mathbf{O}_1[w(s)]$.

Let $t = e^{i\theta}$ denote the values of s on the boundary L of the unit circle S' and allow s to approach t in equation (3.5). In the first place, the functions $\frac{\partial U}{\partial x}$ and $\frac{\partial U}{\partial y}$ must assume the known values on L given by equations (3.2), which now have the form

$$\frac{\partial U}{\partial x} + i \frac{\partial U}{\partial y} = g(\theta) + ih(\theta), \quad \text{on } L, \quad (3.6)$$

where $g + ih$ is given by the right-hand side of equation (3.2) expressed in the co-ordinates of the s -plane. In the second place, since $\mathbf{O}(s)$ and $\mathbf{U}(s)$ are analytic inside L , and since it is desirable to determine these functions from prescribed boundary values given by (3.6), it is convenient to demand that the functions $\mathbf{O}(s)$, $\mathbf{O}'(s)$, and $\mathbf{U}(s)$ approach continuous values on L uniformly as s approaches t from the interior of

† It will be assumed that $w'(s)$ is not zero either in the unit circle S' or on its circumference L . For the region S finite let $w(0) = 0$.

L along any path. Furthermore, it can be shown † that the mapping function $w(s)$ and its derivatives $w'(s)$ and $w''(s)$ approach uniformly to definite limiting values on L whenever the boundary C of the region S is such that the co-ordinates x and y of the contour C have continuous derivatives, with respect to arc-length, to the third order. If these requirements and conditions are satisfied, then it is clear from equation

(3.5) that $\frac{\partial U}{\partial x}$ and $\frac{\partial U}{\partial y}$ must approach uniformly continuous limits on L.

From equation (3.6) the desired results can be obtained if it be assumed that g and h satisfy conditions of the following form on L ‡ :

$$\left. \begin{array}{l} g(\theta) \text{ is continuous, } g'(\theta) \text{ exists, } g \text{ and } g' \text{ satisfy Lipschitz} \\ \text{conditions of the form } |p(\theta) - p(\theta')| \leq h |\theta - \theta'|^m \end{array} \right\}, \quad (3.7)$$

where θ and θ' are two points on L, and h, m are positive constants.

As s approaches t on L equation (3.5) becomes

$$g + ih = O(t) + \frac{w(t)}{\bar{w}'\left(\frac{1}{t}\right)} \bar{O}'\left(\frac{1}{t}\right) + \bar{U}\left(\frac{1}{t}\right). \quad (3.8)$$

Let $p(s)$ be any polynomial which has no zeros inside L and which is such that $p(s)w(s)$ has no poles outside L except perhaps at infinity.

Then multiply the terms in equation (3.8) by $\frac{1}{2\pi i} \frac{p(t) dt}{t-s}$ and integrate the result around the circle L for s interior to L. By use of the Cauchy integral formula and the theory of residues, the function $O(s)$ can be determined. Similarly, the function $U(s)$ can be determined by using the conjugate to equation (3.8).

4. Relations between the Heat Function V and the Biharmonic Function U.

Equation (3.2) gives the relationship between the two functions U and V in that the boundary conditions of U are determined from V. Since only the first derivatives of V occur in this relation (3.2), it is desirable to consider the direct calculation of these derivatives from the

† V. Smirnoff, *Mathematische Annalen*, cvii. pp. 313–323 (1933). In fact Smirnoff proves these results under less restrictive conditions than those given here.

‡ E. Picard, 'D'equations aux derivees partielles' (Paris, 1927), p. 67. It should be noted that if the demands on the functions $O(s)$ and $U(s)$ be less strenuous at the boundary, then the restrictions (3.7) on g and h can be considerably relaxed: see Evans, 'The Logarithmic Potential,' Amer. Math. Soc. Coll. Pub. p. 66 (1927).

differential equation $\nabla^2 V = kT$, where T is assumed to be a continuous function of x and y . A particular integral of this equation is

$$V(x, y) = \frac{k}{2\pi} \int_S T(x_1, y_1) \log p \, dx_1 \, dy_1, \quad (4.1)$$

where (x, y) is a point in S , (x_1, y_1) is a variable point in S , $p^2 = (x - x_1)^2 + (y - y_1)^2$, and S is the area of the cross-section of the body under consideration. If $z = w(s)$ maps the region S conformally on the unit circle of the s -plane, then in terms of the variables of the s -plane the equation (4.1) becomes

$$V = \frac{k}{2\pi} \int_{S'} Tw'(t) \bar{w}'(\bar{t}) \log p \, d\xi_1 \, d\eta_1, \quad (4.2)$$

where $s = \xi + i\eta = re^{i\theta}$, $t = \xi_1 + i\eta_1 = Re^{i\beta}$, S' is the unit circle, and $p^2 = (s - t)(\bar{s} - \bar{t})$; here V has the arguments ξ and η , while T has the arguments ξ_1 and η_1 . Differentiate (4.2) to obtain

$$\frac{\partial V}{\partial x} - i \frac{\partial V}{\partial y} = - \frac{k}{2\pi w'(s)} \int_0^1 \int_0^{2\pi} \frac{Tw'(t) \bar{w}'(\bar{t}) R dR d\beta}{t - s}. \quad (4.3)$$

In order to facilitate calculation it is convenient to express the integral appearing in (4.3) in a form to which the formula of Cauchy and the theory of residues can be applied. If the function $f(R, \beta) \equiv Tw'(t) \bar{w}'(\bar{t})$ is continuous and periodic of period 2π in β and satisfies a Lipschitz condition of the form given in (3.7), then the equation (4.3) can be written in the following form:

$$\frac{\partial V}{\partial x} - i \frac{\partial V}{\partial y} = - \frac{k}{2\pi i w'(s)} \left[\int_0^r + \int_r^1 \right] \int_{|t|=R} \frac{Tw'(t) \bar{w}'(\bar{t}) R dR dt}{t(t-s)}, \quad (4.4)$$

where in the first integral $|t| < |s|$, and in the second integral $|t| > |s|$.

The function $O_1(z)$, defined in equation (2.14) and used in connexion with the biharmonic function U in equation (3.3), will be examined next. The definition of $O_1(z)$ given in (2.14) does not determine $O_1(z)$ completely from $\nabla^2 U$. The function $Q(x, y)$ conjugate to $\nabla^2 U$ is determined except for a real constant, whence $O_1(z)$ is determined except for a term $(a + ib) + Biz$, where a , b , and B are real constants. Since for the calculation of the stresses it is necessary to know only the expression $\nabla^2 U = \sigma_x + \sigma_y + \nabla^2 V$ (see equation (2.12)), it follows that $O_1(z)$ may be subjected to the conditions

$$O_1(0) = 0, \quad \text{Real} \left[\frac{1}{i} O_1'(0) \right] = 0. \quad (4.5)$$

The expressions for the stresses can be readily obtained in terms of the functions $\mathbf{O}_1$, $\mathbf{U}_1$, and V . They are :

$$\left. \begin{aligned} \sigma_x + \sigma_y &= 2[\mathbf{O}_1'(z) + \bar{\mathbf{O}}_1'(\bar{z})] - kT, \\ \sigma_y - \sigma_x + 2i\tau_{xy} &= 2[z\mathbf{O}_1''(z) + \mathbf{U}_1'(z)] - \left[\frac{\partial^2 V}{\partial x^2} - \frac{\partial^2 V}{\partial y^2} - 2i \frac{\partial^2 V}{\partial x \partial y} \right]. \end{aligned} \right\} \quad (4.6)$$

By use of equation (3.3) the equations (2.13) can be combined to give the displacements in the following form :

$$u + iv = \frac{3m-4}{2mG} \mathbf{O}_1(z) - \frac{e_z z}{m} - \frac{1}{2G} \left[z\bar{\mathbf{O}}_1'(\bar{z}) + \bar{\mathbf{U}}_1(\bar{z}) - \left(\frac{\partial V}{\partial x} + i \frac{\partial V}{\partial y} \right) \right]. \quad (4.7)$$

The expressions for the stresses and displacements in terms of the polar co-ordinates r and θ of the s -plane can be obtained by using the following transformations :

$$\left. \begin{aligned} u_r + iv_\theta &= \frac{\bar{s} \bar{w}'(\bar{s})}{r |w'(s)|} (u + iv), & \sigma_r + \sigma_\theta &= \sigma_x + \sigma_y, \\ \sigma_\theta - \sigma_r + 2i\tau_{r\theta} &= \frac{s^2 w'(s)}{r^2 \bar{w}'(\bar{s})} (\sigma_y - \sigma_x + 2i\tau_{xy}). \end{aligned} \right\} \quad (4.8)$$

Now consider the form of the functions $\mathbf{O}_1(z)$ and $\mathbf{U}_1(z)$ in the expressions for the stresses and displacements when the cross-section of the body is multiply-connected. Let the region be bounded by $n+1$ simple closed curves $C_0, C_1, \dots, C_n$, which have no common point and with C_0 enclosing all the other curves. It is permissible for C_0 to recede to infinity. Since the physical considerations demand the stresses, the displacements, and the temperature to be single-valued in S , it is necessary to require that the right-hand members of the equations (4.6) and (4.7) be single-valued. Hence, from the first equation in (4.6), it follows that $\text{Real} [\mathbf{O}_1'(z)] = \frac{1}{4} \nabla^2 U$ must be single-valued. However, in this case, the function $Q(x, y)$ conjugate to $\nabla^2 U$ may not be single-valued and, in fact, may increase by a constant value on going once around a simple closed curve C_k enclosing C_k ($k=1, 2, \dots, n$). Designate this increase by $4A_k$ (A_k real) whence

$$f(z) = \nabla^2 U + iQ(x, y) = \frac{1}{2\pi} \sum_{k=1}^n 4A_k \log(z - z_k) + f^*(z),$$

where $f^*(z)$ is single-valued and analytic in S , and z_k is a point interior to the curve C_k and outside S . Equation (2.14) gives

$$\mathbf{O}_1(z) = \frac{z}{2\pi} \sum_{k=1}^n A_k \log(z - z_k) + \frac{1}{2\pi i} \sum_{k=1}^n (a_k + ib_k) \log(z - z_k) + \mathbf{O}_1^*(z), \quad (4.9)$$

where $\mathbf{O}_1^*(z)$ is single-valued and analytic in S except perhaps at infinity,

and a_k, b_k are real constants. Integration of the terms of equations (4.6), (4.7), and (3.2) around the closed curve C_k' enclosing C_k shows that

$$\left. \begin{aligned} A_k &= 0, \quad a_k - ib_k = \frac{im}{4(m-1)} (X_k - iY_k), \\ [U_1(z)]_{C_k'} &= \left[\frac{\partial V}{\partial x} - i \frac{\partial V}{\partial y} \right]_{C_k'} + i \frac{3m-4}{4(m-1)} (X_k - iY_k), \end{aligned} \right\} \quad (4.10)$$

where
$$X_k + iY_k = - \int_{C_k} (X + iY) ds. \quad (4.11)$$

Thus in a multiply-connected region O_1 and U_1 have the forms

$$\left. \begin{aligned} O_1(z) &= - \frac{m}{8\pi(m-1)} \sum_{k=1}^n (X_k + iY_k) \log(z - z_k) + O_1^*(z), \\ U_1(z) &= \frac{3m-4}{8\pi(m-1)} \sum_{k=1}^n (X_k - iY_k) \log(z - z_k) + U_1^*(z) + M(z), \end{aligned} \right\} \quad (4.12)$$

where $O_1^*(z)$ and $U_1^*(z)$ are single-valued and analytic in S except perhaps at infinity, and $M(z)$ is a multiply-valued function such that

$$[M(z)]_{C_k'} = \left[\frac{\partial V}{\partial x} - i \frac{\partial V}{\partial y} \right]_{C_k'}. \quad (4.13)$$

5. The Long Circular Cylinder of One Material with the Temperature a Function of the Radius.

As an illustration of the application of the method of section 3, let it be required to determine the state of stress in a long circular cylinder of one material in which the temperature is a function of the radius of the cylinder. Assume the temperature T to be a continuous function of the distance r from the centre of the cylinder, and let the radius of the cross-section be b . Suppose the external surface forces to be zero and use the function $z = bs$ to map the cross-section on the unit circle in the s -plane. From equation (4.4) the first derivatives of the heat function V have the following form :

$$\frac{\partial V}{\partial x} - i \frac{\partial V}{\partial y} = \frac{kb}{s} \int_0^r T(bR) R dR. \quad (5.1)$$

Combination of equations (3.6) and (5.1) gives the following relation on the boundary L of the unit circle in the s -plane :

$$g - ih = \frac{M}{t}, \quad \text{with } M = \frac{k}{b} \int_0^b r T(r) dr, \quad (5.2)$$

where $s = t$ on L . Hence the boundary condition (3.8) becomes

$$Mt = O(t) + t\bar{O}'\left(\frac{1}{t}\right) - \bar{U}\left(\frac{1}{t}\right). \quad (5.3)$$

Make use of conditions (4.5) on $\mathbf{O}_1(z)$ and write

$$\mathbf{O}(s) = a_1 s + \sum_{n=2}^{\infty} (a_n + i b_n) s^n, \quad \mathbf{U}(s) = \sum_{n=0}^{\infty} (c_n + i d_n) s^n.$$

Multiply (5.3) by the expression $\frac{1}{2\pi i} \frac{dt}{t-s}$ and integrate over L for s interior to L to obtain

$$Ms = \mathbf{O}(s) + a_1 s + 2(a_2 - i b_2) + (c_0 - i d_0).$$

Take the conjugate of (5.3) and perform the same operations to obtain the following expression :

$$0 = \frac{1}{s} \mathbf{O}'(s) - a_1 \frac{1}{s} + \mathbf{U}(s).$$

From these last two expressions it is clear that $\mathbf{O}(s) = \frac{Ms}{2}$, $\mathbf{U}(s) = 0$, whence $\mathbf{O}_1(z) = \frac{Mz}{2b}$, $\mathbf{U}_1(z) = 0$. It follows from equation (3.3) that $U = \frac{Mr^2}{2b}$, where a constant is neglected since it contributes nothing to the stresses.

Now in polar co-ordinates the classical formulas for the stresses are as follows :

$$\sigma_r = \frac{1}{r} \frac{\partial \mathbf{O}}{\partial r} + \frac{1}{r^2} \frac{\partial^2 \mathbf{O}}{\partial \theta^2}, \quad \sigma_\theta = \frac{\partial^2 \mathbf{O}}{\partial r^2}, \quad \tau_{r\theta} = -\frac{\partial}{\partial r} \left(\frac{1}{r} \frac{\partial \mathbf{O}}{\partial \theta} \right), \quad \dots \quad (5.4)$$

where $\mathbf{O} = U - V$ is the stress function. From the values of U and V obtained above the stresses can be calculated as follows :

$$\left. \begin{aligned} \sigma_r &= \frac{M}{b} - \frac{k}{r^2} \int_0^r r T(r) dr, \quad \tau_{r\theta} = 0, \\ \sigma_\theta &= \frac{M}{b} + \frac{k}{r^2} \int_0^r r T(r) dr - kT, \end{aligned} \right\} \dots \dots \dots (5.5)$$

where M is given by (5.2). Make use of equations (2.4) and (2.7) to obtain $e_z = \frac{2(m-1)M}{Emb}$, $\sigma_z = \frac{2M}{b} - kT$. These results check with those obtained by Timoshenko ⁽⁹⁾ by an entirely different method.

6. The Hollow Circular Cylinder.

It is interesting to extend the methods and results of sections 3 and 4 to the case of the long hollow cylinder. Let S be the region between two closed curves C_0 and C_1 , and let the rational function $z = w(s)$ map the region on a ring formed by two concentric circles about the origin in the s -plane; let C_0 go into circle B of radius b , C_1 into circle A of radius a ($a < b$), and region S into the circular ring between circles A and B .

Now the boundary condition (3.2) must hold on C_0 and C_1 , whence the condition (3.8) must hold on the circles A and B. For this case the equation (3.8) has the following form :

$$\mathbf{O}(t) + \frac{w(t)}{\bar{w}'\left(\frac{a_i^2}{t}\right)} \bar{\mathbf{O}}'\left(\frac{a_i^2}{t}\right) + \bar{\mathbf{U}}\left(\frac{a_i^2}{t}\right) = g_{A_i} + i h_{A_i}, \text{ for } t \text{ on } A_i, i=1, 2. \quad (6.1)$$

Since the region is multiply-connected the functions $\mathbf{O}(s)$ and $\mathbf{U}(s)$ will have a form obtained by transforming the expressions of equations (4.12) to the s -plane. The functions $\mathbf{O}^*(s)$ and $\mathbf{U}^*(s)$ defined in connexion with equations (4.12) will be Laurent series in the ring, so that they can be written in the form

$$\mathbf{O}^*(s) = \mathbf{O}_p^*(s) + \mathbf{O}_d^*(s), \quad \mathbf{U}^*(s) = \mathbf{U}_p^*(s) + \mathbf{U}_d^*(s), \quad \dots \quad (6.2)$$

where $\mathbf{O}_p^*$ and $\mathbf{U}_p^*$ indicate the terms of the Laurent series which have non-negative exponents, while $\mathbf{O}_d^*$ and $\mathbf{U}_d^*$ have negative exponents. In a way analogous to that of section 3 the Cauchy integral formula can be applied to equations (6.1) to obtain the functions $\mathbf{O}_p^*$, $\mathbf{O}_d^*$, $\mathbf{U}_p^*$, $\mathbf{U}_d^*$.

As an illustration, consider the case of a long circular cylinder of radius b with a concentric hole of radius a . Let the temperature T be that arising from a steady state heat distribution so that T is harmonic in the ring S formed by the circles A and B of radii a and b , respectively. Assume that T is expressed in the following series, which converges uniformly in the region S :

$$T = A \log r + A_0 + \sum_{n=1}^{\infty} [(A_n r^n + C_n r^{-n}) \cos n\theta, \quad \left. \begin{aligned} &+ (-B_n r^n + D_n r^{-n}) \sin n\theta \end{aligned} \right\} \dots \quad (6.3)$$

A particular integral of the differential equation $\nabla^2 V = kT$, where T is given by (6.3), is as follows :

$$4V = kAr^2 \log r + k(A_0 - A)r^2 + kr(2 \log r - 1)(C_1 \cos \theta + D_1 \sin \theta) \left. \begin{aligned} &+ 2k[z\bar{G}_1(\bar{z}) + \bar{z}G_1(z) + z\bar{H}_1(\bar{z}) + \bar{z}H_1(z)], \end{aligned} \right\} \quad (6.4)$$

where

$$H_1(z) = - \sum_{n=2}^{\infty} \frac{C_n + iD_n}{4(n-1)} \frac{1}{z^{n-1}}, \quad G_1(z) = \sum_{n=1}^{\infty} \frac{A_n + iB_n}{4(n+1)} z^{n+1}. \quad \dots \quad (6.5)$$

Differentiation of equation (6.4) gives

$$\frac{\partial V}{\partial x} + i \frac{\partial V}{\partial y} = \frac{k}{2} (C_1 + iD_1) \log r + \frac{k}{2} \left(A \log r - \frac{A}{2} + A_0 \right) z \left. \begin{aligned} &+ \frac{k}{4} (C_1 - iD_1) \frac{z}{\bar{z}} + k[z\bar{H}_1'(\bar{z}) + H_1(z) + z\bar{G}_1'(\bar{z}) + G_1(z)], \end{aligned} \right\} \quad (6.6)$$

which is single-valued. Now assume the external surface forces to be

zero and make use of the mapping function $z=s$ to give the following expressions for the right-hand side of (6.1):

$$g_{\Delta_i} + i\hbar_{\Delta_i} = \left[\frac{\partial V}{\partial x} + i \frac{\partial V}{\partial y} \right]_{r=a_i}, \quad i=1, 2. \quad (6.7)$$

In this case the equations (4.12) reduce to $\mathbf{O}_1(z) = \mathbf{O}_1^*(z)$, $\mathbf{U}_1(z) = \mathbf{U}_1^*(z)$. Suppose that the expressions in equation (6.7) satisfy the conditions (3.7) so that the equations (6.1) can be integrated by a method analogous to that outlined in section 3. Use the mapping function $w(s)=s$ and

multiply the equations in (6.1) by $\frac{1}{2\pi i} \frac{dt}{t-s}$. Then integrate the first

equation ($i=1$) of equation (6.1) and its conjugate over circle A for s exterior to A, and integrate the second equation ($i=2$) of (6.1) and its conjugate over B for s interior to B. This gives four equations for determining the functions $\mathbf{O}_p$, $\mathbf{O}_d$, $\mathbf{U}_p$, and $\mathbf{U}_d$ defined in (6.2). The results are:

$$\mathbf{O}_p(s) = kG(s) + \frac{k}{4} \left[\frac{A(b^2 \log b - a^2 \log a)}{b^2 - a^2} - \frac{A}{2} + A_0 \right] s + \frac{k(C_1 - iD_1)s^2}{4(a^2 + b^2)},$$

$$\mathbf{O}_d(s) = kH(s), \quad \mathbf{U}_p(s) = \frac{k}{2} (C_1 - iD_1) \left[\log b - \frac{b^2}{a^2 + b^2} \right],$$

$$\mathbf{U}_d(s) = - \left[\frac{kAa^2b^2}{2(b^2 - a^2)} \log \left(\frac{b}{a} \right) \right] \frac{1}{s} + \left[\frac{ka^2b^2(C_1 + iD_1)}{4(a^2 + b^2)} \right] \frac{1}{s^2},$$

where $G(s) = G_1[w(s)]$, $H(s) = H_1[w(s)]$ (see equation (6.5)). Use $z=w(s)=s$ in these functions to give $\mathbf{O}_1(z)$ and $\mathbf{U}_1(z)$ and then determine $\mathbf{U}$ from equation (3.3). Put the stress function $\mathbf{O} = \mathbf{U} - \mathbf{V}$ in equations (5.4) to obtain the stresses:

$$\left. \begin{aligned} \sigma_r &= \frac{kA}{2} \left[\log \left(\frac{b}{r} \right) - \frac{a^2 b^2 - r^2}{r^2 b^2 - a^2} \log \left(\frac{b}{a} \right) \right] \\ &\quad - \frac{k}{2r^3} \frac{(b^2 - r^2)(r^2 - a^2)}{a^2 + b^2} (C_1 \cos \theta + D_1 \sin \theta), \\ \tau_{r\theta} &= - \frac{k}{2r^3} \frac{(b^2 - r^2)(r^2 - a^2)}{a^2 + b^2} (C_1 \sin \theta - D_1 \cos \theta), \\ \sigma_\theta &= \frac{kA}{2} \left[\log \left(\frac{b}{r} \right) + \frac{a^2 b^2 + r^2}{r^2 b^2 - a^2} \log \left(\frac{b}{a} \right) - 1 \right] \\ &\quad - \frac{k}{2r^3} \frac{[4a^2r^2 - (3r^2 - b^2)(r^2 + a^2)]}{a^2 + b^2} (C_1 \cos \theta + D_1 \sin \theta). \end{aligned} \right\} \quad (6.8)$$

These results check with those that can be obtained from considerations made by Poritsky⁽¹²⁾. However, it should be noted that Poritsky's method applies only for a steady state heat distribution, whereas the method used here can be applied to a quite general distribution of heat.

7. Long Cylindrical Composite Bodies.

The method of solution outlined in sections 3 and 4 can be extended to the case of a long cylindrical body composed of several materials, each of which has different elastic constants E and m and different coefficients of thermal expansion α . Let one material S_0 surround the other n materials $S_1, S_2, \dots, S_n$ in such a way that every material $S_j (j=1, 2, \dots, n)$ is completely surrounded by material S_0 , and designate the boundary of S_j by C_j with the outside boundary being C_0 . In the region S_j the physical constants are E_j, m_j , and α_j .

The materials are assumed to adhere on all junction surfaces so that the displacements will necessarily be continuous on the curves $C_1, \dots, C_n$. A further condition along $C_j (j=1, \dots, n)$ is obtained from the continuity of the normal stress across C_j , while on C_0 the condition (3.2) must hold. Let U_j be the biharmonic function in S_j , V_j the heat function in S_j , $O_{j1}(z)$ and $\bar{U}_{j1}(z)$ the analytic functions in S_j defined by (3.3). Then, in terms of the functions O and $\bar{U}$, the above three conditions can be expressed as follows:

$$O_{01}(z) + z\bar{O}_{01}'(\bar{z}) + \bar{U}_{01}(\bar{z}) = f_{00}(s), \text{ on } C_0, \quad (7.1)$$

$$\left. \begin{aligned} M_j O_{01}(z) + N_j O_{j1}(z) + z\bar{O}_{j1}'(\bar{z}) + \bar{U}_{j1}(\bar{z}) \\ = R_j z + f_{jj}(s), \text{ on } C_j (j=1, 2, \dots, n), \end{aligned} \right\} \quad (7.2)$$

$$\left. \begin{aligned} (N_j - 1)O_{j1}(z) + (M_j + 1)O_{01}(z) + z\bar{O}_{01}'(\bar{z}) + \bar{U}_{01}(\bar{z}) \\ = R_j z + f_{0j}(s), \text{ on } C_j (j=1, 2, \dots, n) \end{aligned} \right\} \quad (7.3)$$

where s is arc-length on C_j , and

$$f_{jj}(s) = \frac{\partial V_j}{\partial x} + i \frac{\partial V_j}{\partial y}, \quad f_{0j} = \frac{\partial V_0}{\partial x} + i \frac{\partial V_0}{\partial y}, \text{ on } C_j, \quad (7.4)$$

$$\left. \begin{aligned} M_j = \frac{4G_j(m_0 - 1)}{m_0(G_0 - G_j)}, \quad N_j = 1 - \frac{4G_0(m_j - 1)}{m_j(G_0 - G_j)}, \\ R_j = -2e_z \frac{G_0 G_j (m_0 - m_j)}{m_0 m_j (G_0 - G_j)}, \quad j=1, 2, \dots, n. \end{aligned} \right\} \quad (7.5)$$

It should be observed that the conditions (7.1), (7.2), and (7.3) become very much simpler when it is assumed that the elastic constants E and m are the same throughout the region $S = S_0 + S_1 + \dots + S_n$, but the coefficient of thermal expansion is different in each region†. Equation (7.1) is unchanged in form, but equations (7.2) and (7.3) reduce to

$$O_{j1}(z) = O_{01}(z), \text{ on } C_j, \quad (7.6)$$

$$\bar{U}_{j1}(\bar{z}) - \bar{U}_{01}(\bar{z}) = f_{jj}(s) - f_{0j}(s), \text{ on } C_j (j=1, \dots, n). \quad (7.7)$$

† This consideration is of some practical interest since there are materials whose elastic properties are nearly alike, but whose coefficients of expansion differ widely.

Hence $O_{j1}(z)$ is the analytic extension of $O_{01}(z)$. Further, if the function V be determined from equation (4.1), then it is known that its first derivatives will be continuous everywhere in the region provided that kT be bounded and integrable. Thus, if V be calculated from (4.1), the right-hand side of equation (7.7) is zero, and $U_{j1}(z)$ is the analytic extension of $U_{01}(z)$. Hence for this case the function U is continuous throughout the region so that the problem is that of the simply-connected region of one material with the boundary condition (7.1). If the region S be mapped conformally on the unit circle by a rational function and if f_{00} satisfies conditions (3.7), then the functions $O_{01}(z)$ and $U_{01}(z)$ can be determined by the method of section 3.

It will now be convenient to examine the general equations (7.1), (7.2), and (7.3) for the case of a body composed of only two materials, one inside the other. Let S_0 be the outside material, C_0 the outside boundary, S_1 the inside material, and C_1 the boundary between the materials. Assume the region to be such that the rational function $z=w(s)$ maps the curves C_0 and C_1 conformally into concentric circles about the origin in the s -plane, with C_1 going into circle A of radius a , and C_0 into circle B of radius b , while S_1 goes into s_1 and S_0 goes into s_0 between the circles A and B . In this case, let $O_{j1}(z) \equiv O_j(s)$, $U_{j1}(z) \equiv U_j(s)$, and $f_{jj} \equiv g_{jj}(s)$ with g_{jj} satisfying conditions (3.7). The functions $O_1(s)$ and $U_1(s)$ in region s_1 can be expressed as power series inside circle A , while the functions $O_0(s)$ and $U_0(s)$ can be expressed as Laurent series in s_0 . Express these latter two functions as in equation (6.2) and use a procedure analogous to that of section 6 to integrate the equations (7.1), (7.2), and (7.3). If the functions are multiply-valued in s_0 , then use equations (4.12).

Consider the case of a long circular cylinder of radius b with a concentric core, of radius a , of a different material. Let B be the outside circular boundary and let A be the circular boundary between the two materials, while S_1 is the material of the core and S_0 is the material of the ring. Take E_i , m_i , α_i , and k_i , where $k_i = \frac{E_i m_i \alpha_i}{m_i - 1}$, as the constants of

material $S_i (i=1, 2)$. Assume the temperature T to be a continuous function of the radius r , and designate by $T_i(r)$ the temperature distribution in $S_i (i=1, 2)$. As is customary, assume the external surface forces to be zero. Use the mapping function $z=w(s)=s$ and determine the first derivatives of V from equation (4.4):

$$\frac{\partial V_1}{\partial x} - i \frac{\partial V_1}{\partial y} = \frac{k_1}{s} \int_0^r r T_1(r) dr, \text{ for } r \leq a,$$

$$= D_1 \bar{t}, \text{ for } t \text{ on } A;$$

$$\frac{\partial V_0}{\partial x} - i \frac{\partial V_0}{\partial y} = \frac{k_1}{s} \int_0^a r T_1(r) dr + \frac{k_0}{s} \int_a^r r T_0(r) dr, \text{ for } a \leq r \leq b, \\ = D_0 \bar{t}, \text{ for } t \text{ on } B,$$

where

$$D_0 = \frac{k_1}{b^2} \int_0^a r T_1(r) dr + \frac{k_0}{b^2} \int_a^b r T_0(r) dr, \quad D_1 = \frac{k_1}{a^2} \int_0^a r T_1(r) dr. \quad (7.8)$$

If these results be substituted in equations (7.1), (7.2), and (7.3), and if these equations be integrated over circles A and B by methods similar to those used in sections 5 and 6, then the following results are obtained for the four functions $O_{01}(z)$, $U_{01}(z)$, $O_{11}(z)$, $U_{11}(z)$:

$$O_{01}(z) = a_1 z, \quad U_{01}(z) = e_1 z^{-1}, \quad O_{11}(z) = c_1 z, \quad U_{11}(z) = 0,$$

where

$$\left. \begin{aligned} a_1 &= \frac{2(R_1 + D_1)a^2 - D_0(N_1 + 1)b^2}{2[M_1 a^2 + (N_1 + 1)(a^2 - b^2)]}, \\ e_1 &= b^2(D_0 - 2a_1), \quad c_1 = \frac{R_1 + D_1 - M_1 a_1}{N_1 + 1}. \end{aligned} \right\} \quad (7.9)$$

From these results the stresses can be calculated:

$$\left. \begin{aligned} \sigma_{r1} &= 2c_1 - \frac{k_1}{r^2} \int_0^r r T_1(r) dr, \quad \tau_{r\theta 1} = 0, \\ \sigma_{\theta 1} &= 2c_1 + \frac{k_1}{r^2} \int_0^r r T_1(r) dr - k_1 T_1, \\ \sigma_{z1} &= E_1 e_z - E_1 \alpha_1 T_1 + \frac{1}{m_1} (4c_1 - k_1 T_1); \end{aligned} \right\} \quad (7.10)$$

$$\left. \begin{aligned} \sigma_{r0} &= 2a_1 + \frac{e_1}{r^2} - \frac{k_1}{r^2} \int_0^a r T_1(r) dr - \frac{k_0}{r^2} \int_a^r r T_0(r) dr, \quad \tau_{r\theta 0} = 0, \\ \sigma_{\theta 0} &= 2a_1 - \frac{e_1}{r^2} + \frac{k_1}{r^2} \int_0^a r T_1(r) dr + \frac{k_0}{r^2} \int_a^r r T_0(r) dr - k_0 T_0, \\ \sigma_{z0} &= E_0 e_z - E_0 \alpha_0 T_0 + \frac{1}{m_0} (4a_1 - k_0 T_0), \end{aligned} \right\} \quad (7.11)$$

where e_z can be determined from the condition

$$\int_0^a \sigma_{z1} r dr + \int_a^b \sigma_{z0} r dr = 0. \quad (7.12)$$

Suppose $T_1 = T_0 = \text{constant} = T$, then the stresses become

$$\left. \begin{aligned} \sigma_{r1} &= \sigma_{\theta 1} = A \left(1 - \frac{b^2}{a^2} \right), \quad \sigma_{r0} = A \left(1 - \frac{b^2}{r^2} \right), \\ \sigma_{\theta 0} &= A \left(1 + \frac{b^2}{r^2} \right), \text{ with } A = \frac{1}{2} (4a_1 - k_0 T). \end{aligned} \right\} \quad (7.13)$$

series developments in s_0 . From equations (6.2) write $\mathbf{O}_0(s) = \mathbf{O}_{0p}(s) + \mathbf{O}_{0d}(s)$ and $\mathbf{U}_0(s) = \mathbf{U}_{0p}(s) + \mathbf{U}_{0d}(s)$. Now integrate the equations (7.1), (7.6), and (7.7) and their conjugates by the method of integration discussed in previous sections to obtain the following results :

$$\left. \begin{aligned} \mathbf{O}_{0d}(s) &= 0, \quad \mathbf{O}_1(s) \equiv \mathbf{O}_{0p}(s) \text{ in } s_1, \\ \mathbf{O}_{0p}(s) &= \frac{T}{4} \left[\left[k_0 + (k_0 - k_1) \frac{a^2 w'(ca^2)}{b^2 w'(cb^2)} \right] w(s) - 2(k_0 - k_1) w\left(\frac{a^2}{b^2} s\right) \right], \\ \mathbf{U}_{0p}(s) &= \frac{T}{2} (k_0 - k_1) \frac{a^2}{b^2} w\left(\frac{b^2}{s}\right) \left[\frac{w'\left(\frac{a^2}{b^2} s\right)}{w'(s)} - \frac{w'(ca^2)}{w'(cb^2)} \right], \\ \mathbf{U}_{0d}(s) &= (k_0 - k_1) T w\left(\frac{a^2}{s}\right), \quad \mathbf{U}_1(s) \equiv \mathbf{U}_{0p}(s) \text{ in } s_1. \end{aligned} \right\} \quad (8.3)$$

By using equations (4.12) and (4.9) the stresses can be calculated in terms of the co-ordinates r and θ of the s -plane :

$$\begin{aligned} \sigma_{r0} + \sigma_{\theta 0} &= (k_0 + k_1) T \frac{a^2}{b^2} \left[\frac{(1 - b^2 c^2)^2}{(1 - a^2 c^2)^2} \right. \\ &\quad \left. - \frac{2b^4 [b^2 - (b^2 + a^2) cr \cos \theta + a^2 c^2 r^2]^2}{(b^4 - 2cra^2 b^2 \cos \theta + a^4 c^2 r^2)^2} \right], \\ \sigma_{r0} - i\tau_{r\theta 0} &= \frac{T}{2} (k_1 - k_0) \frac{a^2}{b^2} \left(1 - \frac{b^2}{r^2} \right) \left[\frac{b^2 (b^2 s^2 - c^2 a^4 r^2) (s - cr^2)^2}{(b^2 s - ca^2 r^2)^2 (s - ca^2)^2} \right. \\ &\quad \left. - \frac{s^2 - c^2 b^2 r^2}{(s - cb^2)^2} \left(\frac{(1 - b^2 c^2)^2}{(1 - a^2 c^2)^2} - \frac{b^4 (1 - cs)^2}{(b^2 - ca^2 s)^2} \right) - \frac{2cb^4 s (b^2 - a^2) (1 - cs) (s - cr^2)}{(s - cb^2) (b^2 - ca^2 s)^3} \right]. \end{aligned}$$

It is clear that if $c=0$ then the mapping function (8.1) is $w(s)=s$, and the stresses should reduce to those for the concentric cylinders of section 7. If one sets $m_1=m_0$, $E_1=E_0$ in equations (7.13) the results obtained are the same as those obtained by setting $c=0$ in the above equations for the stresses.

The stresses in the region s_1 can be calculated in a way similar to those for the region s_0 . The stresses in the co-ordinates of the z -plane can be calculated by transforming the functions of (8.3) to the z -plane by $s = \frac{z}{1+cz}$ and using equations (4.6).

9. Concluding Remarks.

The most important restriction on the theory that has been developed in the preceding sections is that the mapping function be rational. The primary purpose of this restriction has been to provide one with a system of equations that could be solved explicitly for the desired functions from which the stresses in the body could be calculated in explicit form.

On the other hand, if the given contour C , enclosing a simply-connected region S , is sufficiently regular, then it can be mapped on a unit circle to any degree of approximation by a function of the form

$$w(s) = \frac{a}{s} + \sum_{k=0}^n a_k s^k (a=0 \text{ for } S \text{ finite}).$$

In this event, one can readily obtain the solution of the thermo-elastic problem if not for the region S then for some region S' which differs from S as little as desired. Hence, from the practical point of view, the thermo-elastic problem can always be solved by means of such integrals of Cauchy discussed in section 3, the evaluation of which requires only elementary algebraic operations. The situation, naturally, is vastly more involved if the region is multiply-connected.

The assumption concerning the vanishing of the external surface forces made in the foregoing solutions is not essential, since the terms depending on the external surface forces can be carried along without introducing any complications.

Finally, a knowledge of the temperature function T has been assumed throughout the above discussion. But if the temperature follows the law of heat conduction

$$\nabla^2 T = K^2 \frac{\partial T}{\partial t}, \quad (t \text{ is time}),$$

then in many instances it may be possible to solve the stress problem by the foregoing methods when it is not possible to solve this heat conduction equation by elementary methods. For instance, there is no satisfactory and usable solution for the heat conduction equation for the region under consideration in section 8.

References.

- (1) Duhamel, J. M. C., "Memoire sur le calcul des actions moleculaires developpees par les changements de temperature dans les corps solides," *Memoirs . . . de l'institute de France*, v. pp. 440-498 (1838).
- (2) Alibandi, P., "Sulla Elasticita dei Solidi complicata da Variazioni di Temperatura," *Giornale di Matematiche*, xxxviii. pp. 77-91 (1900).
- (3) Leon, A. V., "Spannungen und Formanderungen eines Hohlzylinders und einer Hohlkugel, die von innen erwarmt werden, unter der Annahme eines linearen Temperaturverteilungsgesetzes," *Zeits. f. Math. u. Physik*, lli. pp. 174-190 (1905).
- (4) Lorenz, R., "Temperaturspannungen in Hohlzylindern," *Zeits. des Vereines Deutscher Ingenieure*, li. pp. 743-747 (1907).
- (5) Foppl, A., *Lehrbuch der Technischen Physik*, v. pp. 236-258, Berlin (1907).
- (6) Lees, C. H., "The Thermal Stresses in Solid and in Hollow Circular Cylinders Concentrically Heated," *Proc. Roy. Soc. Math. and Phys.* ci. pp. 411-430, London (1922).

- (7) Dahl, O. G. C., "Temperature and Stress Distributions in Hollow Cylinders," Trans. of the American Society of Mechanical Engineers, xlv. pp. 161-208 (1924).
- (8) Barker, L. H., "The Calculation of Temperature Stresses in Tubes," Engineering, cxxiv. pp. 443-444 (1927).
- (9) Timoshenko, S., 'Theory of Elasticity,' pp. 203-215, 364-380. McGraw-Hill, New York (1934).
- (10) Den Hartog, J. P., "Temperature Stresses in Flat Rectangular Plates and in Thin Cylindrical Tubes," Journal of the Franklin Institute, ccxxii. pp. 149-181 (1936).
- (11) Eichelberg, G., "Temperaturverlauf und Warmespannungen in Verbrennungsmotoren," *Forschungsarbeiten auf den Gebiete des Ingenieurwesens*, Heft 263, 46 pp., Berlin (1923).
- (12) Poritsky, H., "Analysis of Thermal Stresses in Sealed Cylinders and the Effect of Viscous Flow during Anneal," Physics, v. pp. 406-411 (1934); "Thermal Stresses in Cylindrical Pipes," Phil. Mag. ser. 7, xxiv. pp. 209-223 (1937).
- (13) Lebedew, N., "Thermal Stresses in Circular Ring," Applied Mathematics and Mechanics, iii. pp. 76-84 (1936) (in Russian).
- (14) Goodier, J. N., "On the Integration of the Thermoelastic Equations," Phil. Mag. ser. 7, xxiii. pp. 1017-1032 (1937); "Thermal Stress in Long Cylindrical Shells," Canadian Journal of Research, xv. sect. A, pp. 49-58 (1937).
- (15) Biot, M. A., "Propriete General des Tensions Thermiques en Region Stationaire dans Corps Cylindrie," *Societe Scientifiques de Bruxelles*, serie B, liv. pp. 14-18 (1934); "Distributed Gravity and Temperature Loading in Two-dimensional Elasticity replaced by Boundary Pressures and Dislocations," Journal of Applied Mechanics, Trans. A.S.M.E. lvii. pp. A 41 to A 45 (1935); "A General Property of Two-dimensional Thermal Stress Distribution," Phil. Mag. ser. 7, xix. pp. 540-544 (1935).
- (16) Love, A. E. H., 'Mathematical Theory of Elasticity,' Cambridge University Press, London, 4th ed., pp. 49, 131 (1927).
- (17) Muschelisvili, N., "Recherches sur les problemes aux limites relatifs a l'equation biharmonique et aux equations de l'elasticite a deux dimensions," *Mathematische Annalen*, cvii. pp. 282-312 (1933).
- (18) Muschelisvili, N., "Sur l'integration de l'equation biharmonique," *Bulletin de l'Academie des Sciences de Russie*, VI serie, Tome 13, pp. 663-686 (1919); "Praktische Losung der fundamentalen Randwertaufgaben der Elastizitatstheorie in der Ebene fur einige Berandungsformen," *Zeits. f. Angewandte Mathematik und Mechanik*, xiii. pp. 264-282 (1933). See also, I. Malkin, "Uber einige neuere Arbeiten auf den Gebiete der Elastizitatslehre," *Zeits. f. Angewandte Mathematik und Mechanik*, x. pp. 182-197 (1930).
- (19) Muschelisvili, N., *Comptes Rendus de l'Academie des Sciences de l'U.S.S.R.*, Tome III, iii. pp. 7-11, 73-77, 141-144 (1934) (Russian with French summary).

XXIV. On Transient Similarity and Equivalence in n-mesh Linear Networks.

By M. D. COOPER, M.S., E.E., and Professor RICHARD S. BURINGTON,
Ph.D., M.A., Department of Mathematics,
Case School of Applied Science, Cleveland, Ohio †.

[Received September 15, 1939.]

1. Introduction.

IN the present paper the concept of transient similarity is introduced and used to give a complete classification of the set of all n -mesh networks with respect to transient behaviour. The case $n=2$ is considered in detail and gives a complete and improved solution of the problem undertaken by Quade ‡. The connections between steady-state equivalence, transient similarity and transient equivalence are considered in detail.

2. Equivalent Matrices.

Consider a matrix A of order n whose elements are polynomials in λ over the field of real numbers. By means of a sequence Σ of elementary transformations on the rows and columns of A it is possible to reduce A to a diagonal form §

$$N = \begin{pmatrix} \eta_1 & 0 & \dots & 0 \\ 0 & \eta_2 & \dots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \dots & \eta_n \end{pmatrix}, \dots \dots \dots (2.1)$$

in which each η_k divides η_{k+1} , ($k=1, \dots, n-1$). If the rank of A is r , then $\eta_{r+1}=\eta_{r+2}=\dots=\eta_n=0$. Each η_k is a real polynomial in λ and the reduction can be so made that the coefficient of the highest power of λ in each η_k is unity. The sequence Σ defines two real non-singular λ matrices, P and Q , whose determinants are independent of λ , such that

$$N=PAQ. \dots \dots \dots (2.2)$$

† Communicated by the Authors.

‡ Quade, W., 'Klassifikation der Schwingungsvorgänge in gekoppelten Stromkreisen.' Robert Noske in Borna-Leipzig, 1933.

§ Turnbull and Aitken, 'The Theory of Canonical Matrices,' pp. 19-28. Blackie, London (1932).

Two λ matrices A and B of order n over the field F are *equivalent* in F if there exists in F non-singular matrices U and V , whose determinants are independent of λ , such that $B=UAV$. Thus,

Theorem 2.1.—Every matrix A is equivalent to a diagonal matrix N , of the type described in (2.1).

It is known that $d_1=\eta_1$ is the greatest common divisor (g.c.d.) of all the elements of A , and of N ; ...; $d_i=\eta_1\eta_2\ldots\eta_i$ is the g.c.d. of all the i -rowed minor determinants of A and of N ; ... ($i=1, \ldots, n-1$). Furthermore, each d_i divides d_{i+1} . N is called the *Smith normal form* of A .

The quotients $\eta_1=d_1$, $\eta_2=d_2/d_1, \ldots, \eta_i=d_i/d_{i-1}, \ldots$ are called *invariant factors* of A . The invariant factors of A are invariants under the relation of equivalence, and are uniquely determined within a constant factor. Hence the normal form N of A is an invariant for all matrices equivalent to A .

From (2.2), the determinant of N is $d(N)=d(P).d(A).d(Q)$. Since $d(P)$ and $d(Q)$ are non-zero and independent of λ , the roots of the equations $d(N)=0$ and $d(A)=0$ are identically the same. The equation $d(N)=0$ is called the *determinantal equation* of A . Thus the roots of the determinantal equation are absolute invariants of A under equivalence.

Suppose each η_i in N is expressed as a product of powers of irreducible real factors $\dagger$,

$$\eta_i=w_1^{f_{i1}}w_2^{f_{i2}}\ldots w_k^{f_{ik}}v_1^{g_{i1}}v_2^{g_{i2}}\ldots v_l^{g_{il}}, \quad (i=1, \ldots, n), \quad (2.3)$$

where $w_j=\lambda-r_j$; $v_j=(\lambda^2+a_j\lambda+b_j)$;

and where $r_1, \ldots, r_k, a_j$ and b_j are all real, and v_j is irreducible in the real field; k is the number of distinct real roots and $2l$ the number of distinct complex roots of $d(N)=0$. Since η_i divides η_{i+1} , the exponents of each factor form a sequence

$$f_{nj} \geq f_{n-1,j} \geq \ldots \geq f_{1,j}; \quad g_{nu} \geq g_{n-1,u} \geq \ldots \geq g_{1,u}, \quad (j=1, \ldots, k; \quad u=1, \ldots, l). \quad (2.4)$$

Such of these powers $w_s^{f_{is}}, v_t^{g_{it}}$ as are not constants are called the *real elementary divisors* of A . They are defined within a constant factor. It is known that

Theorem 2.2.—A matrix B is equivalent to a matrix A if and only if B and A have the same elementary divisors (invariant factors).

The invariant factors determine the elementary divisors, and conversely.

Associated with the group of all matrices equivalent to A , there are two types of invariant: (1) the distinct roots $r_1, \ldots, r_h, h=k+2l$, of the

$\dagger$ C. C. MacDuffee, 'Theory of Matrices,' *Erg. der Math.* p. 43, Berlin (1932).

determinantal equation, and (2) the exponents $\dots, f_{ij}, \dots; \dots, g_{iu}, \dots$ of the elementary divisors, called *invariants of structure*.

We shall write these invariants of structure

$$[(f_{n1}f_{n-1,1} \dots)(f_{n2}f_{n-1,2} \dots) \dots; (g_{n1}g_{n-1,1} \dots)(g_{n2}g_{n-1,2} \dots) \dots].$$

This symbol is known as the *Segre characteristic* of the matrix A. When the rank r of A is less than n , we shall prefix the Segre characteristic for A by $*^{n-r}$; e. g., $[*^{n-r}(f_{r1}f_{r-1,1} \dots) \dots; \dots]$.

3. Linear Differential Equations.

Consider the system of n linear differential equations with constant coefficients

$$\sum_{s=1}^n a_{rs} q_s = E_r, \quad (r=1, \dots, n) \quad (3.1)$$

where each a_{rs} is a polynomial in λ over the real field, $\lambda = d/dt$, and where each q_s and E_r is a real function of the real independent variable t .

We write (3.1) in the matrix form

$$A\{q\} = \{E\}, \quad (3.2)$$

where $A \equiv (a_{rs})$.

Suppose that the Smith normal form of A is N, and that

$$N = PAQ, \quad (3.3)$$

where P and Q are each non-singular and have determinants which are independent of λ . From (3.1)

$$PA\{q\} = PAQQ^{-1}\{q\} = P\{E\} = \{F\}. \quad (3.4)$$

Let

$$Q^{-1}\{q\} = \{x\}. \quad (3.5)$$

Then, by (3.3) and (3.4),

$$N\{x\} = \{F\}. \quad (3.6)$$

This is a set of differential equations in which the variables x_i of $\{x\}$ are all isolated. Hence, each equation of (3.6) can be solved independently of the rest by the usual elementary methods. Suppose the general solution of (3.6) is

$$\{x\} = \{s\}. \quad (3.7)$$

Then from (3.5),

$$\{q\} = Q\{s\}, \quad (3.8)$$

is the general solution of the original set of equations (3.2).

We shall be concerned principally with the complementary solution of (3.2), that is, the solution of

$$A\{q\} = \{0\}, \quad (3.9)$$

with the corresponding normal equation

$$N\{x\} = \{0\}. \quad (3.10)$$

Let the i th equation in (3.10) be

$$\eta_i x_i = 0, \quad \dots \quad (3.11)$$

where η_i is a polynomial of degree k_i in λ . Write η_i in the factored form

$$\eta_i = (\lambda - r_{i1})(\lambda - r_{i2}) \dots (\lambda - r_{ik_i}). \quad \dots \quad (3.12)$$

It is well known that if the roots (real and complex) $r_{i1}, r_{i2}, \dots, r_{ik_i}$ are all distinct, the general solution of (3.11) is

$$x_i = \sum_{j=1}^{k_i} c_{ij} e^{r_{ij} t}, \quad \dots \quad (3.13)$$

where the c_{ij} are arbitrary (constants of integration). If, however, a root r_{ij} is of multiplicity u , then in order that the solution remain general, the terms in (3.13) involving r_{ij} must be of the form

$$(c_{ij} + c_{i,j+1}t + \dots + c_{i,j+u-1}t^{u-1})e^{r_{ij}t}. \quad \dots \quad (3.14)$$

Each equation in (3.10) has a general solution of the form given in (3.13), subject, in the case of each multiple root, to modifications of type (3.14). It follows from (3.8) that each q_i in $\{q\}$ of (3.9) is a linear function of terms of the type appearing in (3.13) (and in the case of multiple roots, modified as indicated in (3.14)). In fact, in case all the roots $r_1, r_2, \dots, r_h$ of the determinantal equation of N are simple, $e^{r_1 t}, \dots, e^{r_h t}$ form a fundamental set $\dagger$ of solutions for (3.9). Suppose, however, r_j is of multiplicity u_j , $j=1, \dots, h$. Then

$e^{r_1 t}, te^{r_1 t}, \dots, t^{u_1-1}e^{r_1 t}, \dots, e^{r_j t}, te^{r_j t}, \dots, t^{u_j-1}e^{r_j t}, \dots, e^{r_h t}, te^{r_h t}, \dots, t^{u_h-1}e^{r_h t}$

form a fundamental set of solutions for (3.9).

The total number of independent constants of integration in the general solution of (3.2) is the degree in λ of $d(N)$. The degree of $d(N)$ is equal to the degree of $d(A)$.

4. Linear Networks $\ddagger$.

We consider an n -mesh m -terminal pair linear electrical network $\mathcal{N}$.

Let $e_1, \dots, e_m$ denote the (real) external e.m.f.'s applied to the terminal pairs $1, \dots, m$, respectively; q_r and i_r the (real) charge and current, respectively, in mesh r ; $\frac{dq_r}{dt} = i_r$, $q_r = \int_{-\infty}^t i_r dt$; L_{rs} , R_{rs} , and S_{rs} the (real) inductance, resistance, and elastance common to meshes r and s ; L_{rr} , R_{rr} , and S_{rr} the total (real) inductance, resistance, and elastance in mesh r .

$\dagger$ Moulton, F. R., 'Differential Equations,' pp. 246-290.

$\ddagger$ Burington, R. S., "Matrices in Electric Circuit Theory," pp. 325-349, J. Math. and Physics (Dec. 1935).

where $L \equiv (L_{rs})$, $R \equiv (R_{rs})$, and $S \equiv (S_{rs})$ are, respectively, the inductance, resistance, and elastance matrices (all real, and of order 2) of $\mathcal{N}$. Let the determinant of A be

$$\Delta = \alpha_0 \lambda^4 + \alpha_1 \lambda^3 + \alpha_2 \lambda^2 + \alpha_3 \lambda + \alpha_4. \quad (6.2)$$

Suppose the Smith normal form of A is

$$N = \begin{pmatrix} \eta_1 & 0 \\ 0 & \eta_2 \end{pmatrix}. \quad (6.3)$$

Case A. L non-singular.—Since $d(L) = \alpha_0$, Δ is of degree 4 when L is of rank 2. Since η_1 is the (real) g.c.d. of all the elements a_{rs} of A , it follows that the degree of η_1 cannot exceed 2. As indicated in Table I. there are three possible classes to consider, according as the real g.c.d. η_1 of the elements of A is of degree 0, 1, or 2.

TABLE I.

Degree of	Class I.	Class II.	Class III.
η_1	0	1	2
η_2	4	3	2

Classes I., II., III. are each divided into subclasses N in which no root of $\Delta = 0$ is zero, and subclasses Z in which $\Delta = 0$ has one or more zero roots. Since $d(S) = \alpha_4$, no root of $\Delta = 0$ is zero when S is of rank 2. If S is singular, $\Delta = 0$ has one or more zero roots.

The subclasses N and Z are each further divided according to the Segre characteristic (MS), as indicated in Table II. The classification used by Quade is indicated in the column marked *Seg.* The distinctions indicated in MS were not made by Quade.

Cases B, C, D, E. L singular.—When L is singular, Δ is of degree less than four. These cases are clearly indicated in Table II., where h is the degree of Δ . If matrix A is singular, $\eta_2 = 0$, and η_1 is of degree 2, 1, 0, or vanishes completely. The case $\eta_1 = \eta_2 = 0$ is of no interest.

7. Physical Interpretation of Classification.

The Segre characteristic for a given network $\mathcal{N}$ of matrix A , as described in §6, clearly indicates the type and number of transient modes present in the transient solution of $\mathcal{N}$. Thus, the transient solution of any network of characteristic [1111;] involves four distinct terms (modes), all aperiodic; of characteristic [; 11], involves oscillatory modes of two distinct frequencies; of characteristic [11; 1], involves two aperiodic modes and oscillatory terms of one frequency.

TABLE II.
Case A ($h=4$).

Class I-N.		Class I-Z.	
Seg.	MS.	Seg.	MS.
[1111]	[1111 ;] [; 11] [11 ; 1]	[111̇]	[111̇ ;] [1̇ ; 1] [21̇ ;]
[211]	[211 ;] [2 ; 1]	[211̇]	[211̇ ;] [2̇ ; 1]
[22]	[22 ;] [; 2]	[22̇]	[22̇ ;] [31̇ ;]
[31]	[31 ;]	[31̇]**	[31̇ ;]**
[4]	[4 ;]	[4̇]**	[4̇ ;]**

Case B ($h=3$).

[111]	[111 ;] [1 ; 1]	[111̇]	[111̇ ;] [1̇ ; 1]
[21]	[21 ;]	[21̇] [2̇]	[21̇ ;] [2̇ ;]
[3]	[3 ;]	[3̇]*	[3̇ ;]*

Case C ($h=2$).

[11]	[11 ;] [; 1]	[11̇]	[11̇ ;] [1̇ ;]
[2]*	[2 ;]*	[2̇]*	[2̇ ;]*

Case D ($h=1$).

[1]	[1 ;]	[1̇]	[1̇ ;]
-----	-------	------	--------

Case E ($h=0$).

[0]	[;] [* ;]	Vacuous	
-----	---------------	---------	--

TABLE II. (continued).

Case A (*h*=4).

Class II-N.		Class II-Z.	
$[(11)11]$	$[(11)11 ;]$ $[(11) ; 1]$	$[(11)\overset{\circ}{1}\overset{\circ}{1}]$ $[(\overset{\circ}{1}\overset{\circ}{1})11]$	$[(11)\overset{\circ}{1}\overset{\circ}{1} ;]$ $[(\overset{\circ}{1}\overset{\circ}{1})11 ;]$ $[(\overset{\circ}{1}\overset{\circ}{1}) ; 1]$
$[(11)2]**$	$[(11)2 ;]**$	$[(11)\overset{\circ}{2}]^*$ $[(\overset{\circ}{1}\overset{\circ}{1})2]**$	$[(11)\overset{\circ}{2} ;]^*$ $[(\overset{\circ}{1}\overset{\circ}{1})2 ;]**$
$[(21)1]$	$[(21)1 ;]$	$[(21)\overset{\circ}{1}]$ $[(\overset{\circ}{2}\overset{\circ}{1})1]$	$[(21)\overset{\circ}{1} ;]$ $[(\overset{\circ}{2}\overset{\circ}{1})1 ;]$
$[(31)]**$	$[(31) ;]**$	$[(\overset{\circ}{3}\overset{\circ}{1})]**$	$[(\overset{\circ}{3}\overset{\circ}{1}) ;]**$

Case B (*h*=3).

$[(11)1]$	$[(11)1 ;]$	$[(11)\overset{\circ}{1}]$ $[(\overset{\circ}{1}\overset{\circ}{1})1]$	$[(11)\overset{\circ}{1} ;]$ $[(\overset{\circ}{1}\overset{\circ}{1})1 ;]$
$[(21)]$	$[(21) ;]$	$[(\overset{\circ}{2}\overset{\circ}{1})]$	$[(\overset{\circ}{2}\overset{\circ}{1}) ;]$

Case C (*h*=2).

$[(11)]$	$[(11) ;]$	$[(\overset{\circ}{1}\overset{\circ}{1})]$	$[(\overset{\circ}{1}\overset{\circ}{1}) ;]$
----------	-------------	--	---

Case D (*h*=1).

$[(1)]$	$[*1 ;]$	$[(\overset{\circ}{1}) ;]$	$[*\overset{\circ}{1} ;]$
---------	-----------	-----------------------------	----------------------------

Case E vacuous.

Class III-N.	Class III-Z.
--------------	--------------

Case A (*h*=4).

$[(11)(11)]$	$[(11)(11) ;]$ $[; (11)]$	$[(11)(\overset{\circ}{1}\overset{\circ}{1})]$ $\overset{\circ\circ}{[(22)]}$	$[(11)(\overset{\circ}{1}\overset{\circ}{1}) ;]$ $\overset{\circ\circ}{[(22) ;]}$
--------------	--------------------------------	---	---

TABLE II. (continued).

Case C ($h=2$).

$[*11;]$	$[\overset{\circ}{*}1\overset{\circ}{1};]$
$[*2;]$	$[\overset{\circ}{*}2;]$
$[*; 1]$	

Cases B, D, E vacuous.

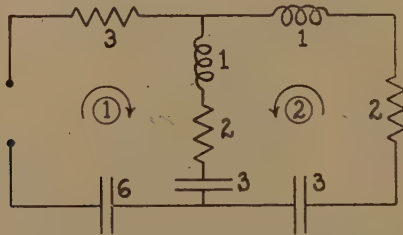
Example.—Consider the network $\mathcal{N}$ of matrix

$$A \equiv \begin{pmatrix} \lambda^2 + 5\lambda + 9, & \lambda^2 + 2\lambda + 3 \\ \lambda^2 + 2\lambda + 3, & 2\lambda^2 + 4\lambda + 6 \end{pmatrix} \quad (7.1)$$

and equation

$$A\{q\} = \{0\}. \quad (7.2)$$

Fig. 1.



By elementary transformations the normal form of (7.1) is found to be

$$N \equiv \begin{pmatrix} 1 & 0 \\ 0 & \Delta \end{pmatrix} = PAQ, \quad (7.3)$$

where

$$\Delta = (\lambda + 3)(\lambda + 5)(\lambda^2 + 2\lambda + 3),$$

$$P \equiv \begin{pmatrix} \frac{3\lambda^2 + 5\lambda + 9}{9} & \frac{-3\lambda^2 - 14\lambda - 24}{9} \\ -\lambda^2 - 2\lambda - 3 & \lambda^2 + 5\lambda + 9 \end{pmatrix}, \quad Q \equiv \begin{pmatrix} 1 & M \\ 0 & 1 \end{pmatrix}, \quad (7.4)$$

and

$$9M = +3\lambda^4 + 29\lambda^3 + 94\lambda^2 + 147\lambda + 117.$$

The characteristic of A is $[11; 1]$, which indicates that the transients of N involve two aperiodic modes and oscillatory terms of one frequency.

The solutions of $N\{x\} = \{0\}$ are (see § 3)

$$\left. \begin{aligned} x_1 &= 0, \\ x_2 &= c_1 e^{-3t} + c_2 e^{-5t} + e^{-t} [c_3 \cos \sqrt{2}t + c_4 \sin \sqrt{2}t]. \end{aligned} \right\} \quad (7.5)$$

Then from (3.5), $\{q\} = Q\{x\}$ gives the transient solutions of (7.2), hence

$$q_1 = -2c_1 e^{-3t} - 2c_2 e^{-5t}, \quad q_2 = x_2. \quad (7.6)$$

From §3 we know that the charge (and current) in mesh k , ($k=1, \dots, n$), of $\mathcal{N}_1$ is a linear function $\mathcal{L}_{1k}$ of $e^{r_1 t}, \dots, e^{r_u t}, \dots, t e^{r_u t}, \dots$, where $r_1, \dots, r_u, \dots$, are the zeros of the determinantal equation of $\mathcal{N}_1$.

These linear functions $\mathcal{L}_{1k}$ are functions of the geometric character (the topological graph) of $\mathcal{N}_1$ as well as the values and arrangement of the circuit parameters.

Consider two networks $\mathcal{N}_1$ and $\mathcal{N}_2$ of matrices A_1 and A_2 , having normal forms N_1 and N_2 , respectively. If $\mathcal{N}_1$ is transiently equivalent to $\mathcal{N}_2$, then $N_1 = N_2$. Suppose $N_1 = P_1 A_1 Q_1$ and $N_2 = P_2 A_2 Q_2$, $N_1 = N_2$. Let the solution of $N_1\{x\} = \{0\}$ be $\{x\} = \{s\}$. Then $\mathcal{L}_{1k}$ is the k th row of $Q_1\{s\}$ and $\mathcal{L}_{2k}$ is the k th row of $Q_2\{s\}$. Thus, while $\mathcal{N}_1$ and $\mathcal{N}_2$ are transiently equivalent, their linear functions $\mathcal{L}_{1k}$ and $\mathcal{L}_{2k}$ may be different, due to the absence of one or more modes in either $\mathcal{L}_{1k}$ or $\mathcal{L}_{2k}$. Furthermore, even though $\mathcal{L}_{1k}$ and $\mathcal{L}_{2k}$ contain exactly the same modes, the actual transient response of $\mathcal{N}_1$ and $\mathcal{N}_2$ may be different in view of the fact that their constants of integration differ, these constants depending on the initial conditions of charges and currents in the respective networks.

Thus *transient equivalence does not imply absolute transient identity*.

If a network $\mathcal{N}$ of matrix $A = L\lambda^2 + R\lambda + S$ is passive then L , R , and S must each be positive semi-definite, and one of them must be positive definite. In many cases there exists no passive network of matrix N , where N is the Smith normal form of A . Furthermore, in many cases there exists no passive network $\mathcal{N}_1$ of matrix A_1 , where A_1 is equivalent to A .

To meet this difficulty Quade introduced a matrix which he called a *vertreter*. A *vertreter* V of a normal matrix N is a real λ matrix satisfying the conditions: (a) V is equivalent to N ; (b) the elements of V are polynomials in λ of degree two or less; (c) $V = V_0\lambda^2 + V_1\lambda + V_2$, where V_0, V_1, V_2 are each symmetric and positive semi-definite †.

The conditions for the existence of a *vertreter* to any given N are not known except for certain special cases, nor does N uniquely determine a V . It was Quade's idea that if a V could be found for a given N , then perhaps a physically realizable passive network could be completely specified whose matrix is V . But unfortunately, even if a network *graph* be found for V , one has no assurance that the resulting theoretical network will be actually physically realizable, for realization of a network having matrix V may require the use of negative resistance, ideal transformers, or the like.

† Quade required that V_0, V_1, V_2 each be positive definite.

The results of this paper appear to be a distinct improvement on the earlier work of Quade. Quade's classification (column Seg. in Table II.) depends on the Segre characteristic for a complex field, and hence failed to distinguish clearly between the oscillatory and aperiodic cases. It is felt that a distinctly better separation from both the physical and the mathematical points of view is given in the column marked MS.

The authors have found 2-mesh passive realizable networks corresponding to each of the characteristics listed in Table II., except as marked by the asterisk (*).

Characteristics marked (**) have no Vertreter.

8. *Extension of Classification to n -mesh Networks.*

A classification of the set of all n -mesh networks, having network matrices A of order n , with respect to their transient behaviour, may be made according to the Segre characteristic of A , following the same scheme illustrated in §6 for the 2-mesh case. This classification has been carried out in detail † for the case $n=3$.

9. *Transient Similarity and Steady-state Equivalence.*

In connexion with the study of equivalent networks one may ask whether steady-state equivalence implies transient similarity (or equivalence), and conversely. We recall the definition of (steady-state) equivalent networks.

Consider the m -terminal pair network $\mathcal{N}$ discussed in §4. Let $B \equiv (b_{rs})$ be the matrix A with λ considered as an indeterminate. (Often in electrical work $\lambda = j\omega$, $j^2 = -1$, where $\omega/2\pi$ is the frequency of the impressed e.m.f.) Suppose B is non-singular. Let $Y \equiv (Y_{rs})$ denote B^{-1} with all but the first m rows and the first m columns deleted. Y is called the *characteristic coefficient matrix* of $\mathcal{N}$.

Two m -terminal pair linear networks $\mathcal{N}_1$ and $\mathcal{N}_2$ having characteristic coefficient matrices Y_1 and Y_2 are said to be *steady-state equivalent* if Y_1 is identically equal to Y_2 in λ .

Let $\mathcal{N}_1$ be a network of n meshes and m -terminal pairs and having the network matrix A_1 of order n . Let P be any m -affine ‡, real and non-singular matrix. It is known that all m -terminal pair networks $\mathcal{N}$ of matrix $A = P^T A_1 P$, are steady-state equivalent to $\mathcal{N}_1$. (In both $\mathcal{N}_1$ and $\mathcal{N}$, the meshes through the terminal pairs are numbered $1, \dots, m$; and in every case, but one mesh passes through a given set of terminal pairs.) Since A is congruent to A_1 , A is equivalent to A_1 , so that in

† M. D. Cooper, M.S. 'Thesis,' Case School of Applied Science, 1939.

‡ R. S. Burington, *loc. cit.* p. 342.

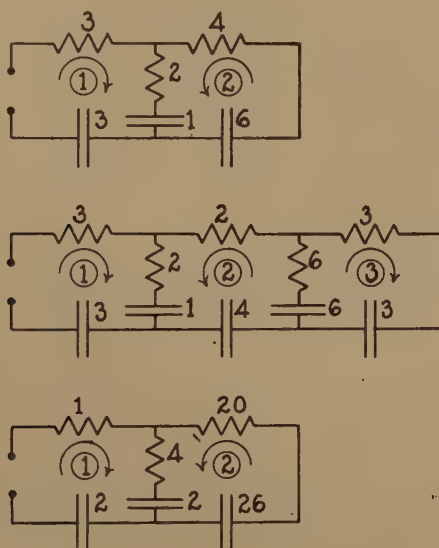
addition to being steady-state equivalent, $\mathcal{N}_1$ and all $\mathcal{N}$ are transiently equivalent, and hence transiently similar.

It is easy to show that steady-state equivalence does not necessarily imply transient equivalence, or even similarity. Thus, the networks $\mathcal{N}_1$ and $\mathcal{N}_2$ of matrices A_1 and A_2 , where

$$A_1 \equiv \begin{pmatrix} 5\lambda+4, & 2\lambda+1 \\ 2\lambda+1, & 6\lambda+7 \end{pmatrix}, \quad A_2 \equiv \begin{pmatrix} 5\lambda+4, & 2\lambda+1, & 0 \\ 2\lambda+1, & 10\lambda+11, & 6\lambda+6 \\ 0, & 6\lambda+6, & 9\lambda+9 \end{pmatrix},$$

are steady-state equivalent but not transiently similar, for their normal forms are, respectively,

Fig. 2.



$$N_1 \equiv \begin{pmatrix} 1 & 0 \\ 1 & \Delta_1 \end{pmatrix}, \quad N_2 \equiv \begin{pmatrix} 1, & 0, & 0 \\ 0, & 1, & 0 \\ 0, & 0, & 9(\lambda+1)\Delta_1 \end{pmatrix},$$

their characteristics are $[11;]$ and $[111;]$, respectively, and their characteristic coefficient matrices are both $\left(\frac{6\lambda+7}{\Delta_1}\right)$, where $\Delta_1 = 26\lambda^2 + 55\lambda + 27$.

However, the networks $\dagger \mathcal{N}_1$ and $\mathcal{N}_3$ of matrices A_1 and A_3 , where

$$A_1 \equiv \begin{pmatrix} 5\lambda+4, & 2\lambda+1 \\ 2\lambda+1, & 6\lambda+7 \end{pmatrix}, \quad A_3 \equiv \begin{pmatrix} 5\lambda+4, & 4\lambda+2 \\ 4\lambda+2, & 24\lambda+28 \end{pmatrix},$$

$\dagger$ The choice of parameters in the above illustrations has been made solely for mathematical simplicity.

are steady-state equivalent and also transiently equivalent for A_3 , are equivalent to A_1 , and their characteristic coefficient matrices are

$$Y = \left(\frac{6\lambda + 7}{\Delta_1} \right).$$

All networks $\mathcal{N}$, of matrix $A = Q^T A_1 Q$, where Q is any non-singular λ matrix whose determinant is independent of λ , are transiently equivalent to $\mathcal{N}_1$. However, as is well known, $\mathcal{N}$ is not necessarily steady-state equivalent to $\mathcal{N}_1$.

XXV. *Hyperfine Structure in the Arc Spectrum of Tin.*

By S. TOLANSKY, Ph.D., and G. O. FORESTER, Ph.D.,
The University, Manchester *.

[Received April 28, 1941.]

ABSTRACT.

Hyperfine structure measurements are made on lines of the tin arc spectrum, most of which lie in the ultra-violet region. It is confirmed that the two odd isotopes 117, 119, have nuclear spins of $\frac{1}{2}$ and identical nuclear magnetic moments. No even isotope displacements are observed, and from observations upon reversals it is shown that any existing even isotope displacements must be very small. The classified lines amongst those observed are analysed, and hyperfine structure interval factors are derived for seven terms. Amongst these are the complete group belonging to the deep $5p^2$ electron configuration. The nuclear coupling of the $5p^2$ electron group is shown to be large, in agreement with deductions about the nuclear coupling of the $5p^4$ electron group in iodine. A J-value has been allocated to one term on the basis of hyperfine structure evidence.

Introduction.

The arc spectrum of tin is difficult to excite under conditions suited to the examination of the hyperfine structures of the lines. The structure in only one line has so far been detected (Tolansky, 1934). Tin has a complex isotopic constitution, the percentage abundances of the isotopes being shown in Table I.

The two main odd isotopes, 117, 119, together make up almost the whole odd isotopic contribution, the intensities of the remaining two being too small to have any effect at all upon the spectroscopic observations. Tolansky (1933, 1934) has already proved, from a study of lines in the spark spectrum, that the two main odd isotopes both have the same nuclear spin equal to $\frac{1}{2}$, and, within the limits of observation, the same nuclear magnetic moment, -0.89 proton magnetons.

As in the spark lines, no even isotope displacement effect occurs in the arc lines, hence each line pattern consists of a single strong even isotope

* Communicated by S. Tolansky.

component comprising 76 per cent. of the total intensity, the remaining 24 per cent. being distributed amongst the individual components of the hyperfine structure multiplet of the odd isotopes. Even although all the odd isotope corresponding components are superposed, because of the identical spins and magnetic moments, this leads to difficulties in analysis. Thus, in a typical $\Delta J=1 \rightarrow 2$ transition, the intensities of the spin multiplet components are $14\frac{1}{2}$, 8, $1\frac{1}{2}$ per cent. respectively. In the presence of the 76 per cent. even isotope component it is impossible to detect the weakest of these, and indeed it so happens that the overall intensity of some of the lines is insufficient to permit of high accuracy in estimating the position of the 8 per cent. component. This is due partly to the strength of the even isotope line, but mostly because of our

TABLE I.

Abundance (per cent.).	Even isotopes.	Odd isotopes.	Abundance (per cent.).
1.1	112		
0.7	114		
		115	0.4
14.2	116		
		117	9.8
21.5	118		
		119	11.0
27.0	120		
		121	2.9
5.0	122		
6.2	124		
Total	75.7	Total	24.3

being unable to excite the tin arc spectrum free from impurity background, despite the employment of many different types of source under very varying conditions of excitation.

In the present investigations measurements have been made upon eight lines, six of which have been classified. All but one of the lines lie in the ultra-violet region, hence for the purpose of analysis we have used a quartz Lummer plate and an aluminized quartz Fabry-Perot interferometer. Tolansky's earlier conclusions about the nuclear spins are confirmed. Hyperfine structure interval factors are derived for a number of terms.

Experimental.

We have found it difficult to excite the spectrum of Sn 1 and yet retain conditions of temperature, pressure and reversal suitable for hyperfine

structure observations. We have experimented with many different types of source and in no case could improve upon a hot hollow cathode discharge. The currents permissible did not result in a very bright spectrum, and we invariably found the lines accompanied by a background of impurity spectrum which added to the difficulties of observation, reducing the accuracy in determining the positions of the weak com-

TABLE II.

Analysis.

Wave-length.	Allocation.	Structure $\text{cm.}^{-1} \times 10^{-3}$.		
		Red.	Violet.	
5632	$5p^2 {}^1S_0 - 5p6s {}^3P_1$	81 (2)	0 (10)	160 (1)
3352		294 (1)	0 (10)	216 (2½)
3223	$5p^2 {}^1D_2 - 19529$	220 (2)	0 (10)	374 (1½)
*3034	$5p^2 {}^3P_1 - 5p6s {}^3P_0$			
3009	$5p^2 {}^3P_1 - 5p6s {}^3P_1$	136 (2)	0 (10)	200 (1)
*2863	$5p^2 {}^3P_0 - 5p6s {}^3P_1$		0 (10)	166 (—)
2832		313 (1)	0 (10)	120 (1½)
2762	$5p^2 {}^3P_2 - 19529$	236 (2)	0 (10)	387 (1)

* Reversed. See discussion.

ponents. Since our main object is the evaluation of the hyperfine structure interval factors of the terms involved, and as this analysis is not concerned with perturbation, deviations from the interval rule or isotope displacement effects, this reduction of accuracy is of no serious consequence.

For the resolving of the structures we have employed a quartz Lummer plate by Hilger, 20 cm. long and 3.42 mm. thick, and to confirm the order positions a 7 cm. diameter quartz Fabry-Perot etalon of variable gap,

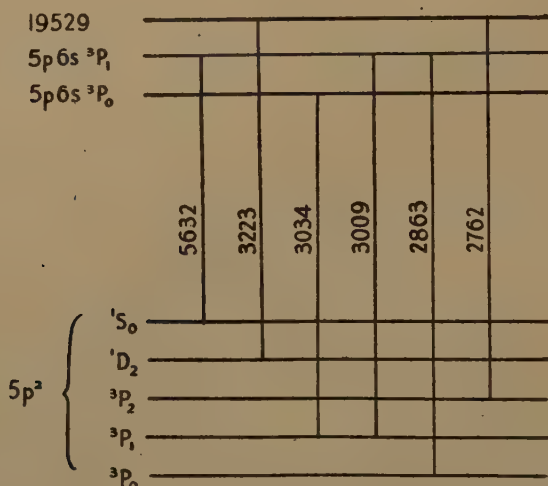
coated with aluminium by evaporation *in vacuo*. Both interferometers were crossed with a Hilger medium quartz spectrograph.

The observed hyperfine structures are shown in Table II., the displacements of the spin components from the even isotope line being given in $\text{cm.}^{-1} \times 10^{-3}$. Under each component the visually estimated intensity is shown in brackets. The line classifications used are those given recently by Meggers (1940).

Analysis.

The multiplet transitions under discussion are shown in fig. 1. Neither the *L* nor the *J* value has yet been determined for the term at 19529 cm.^{-1} . We shall show later that it is highly probable that the *J* value is 1.

Fig. 1.



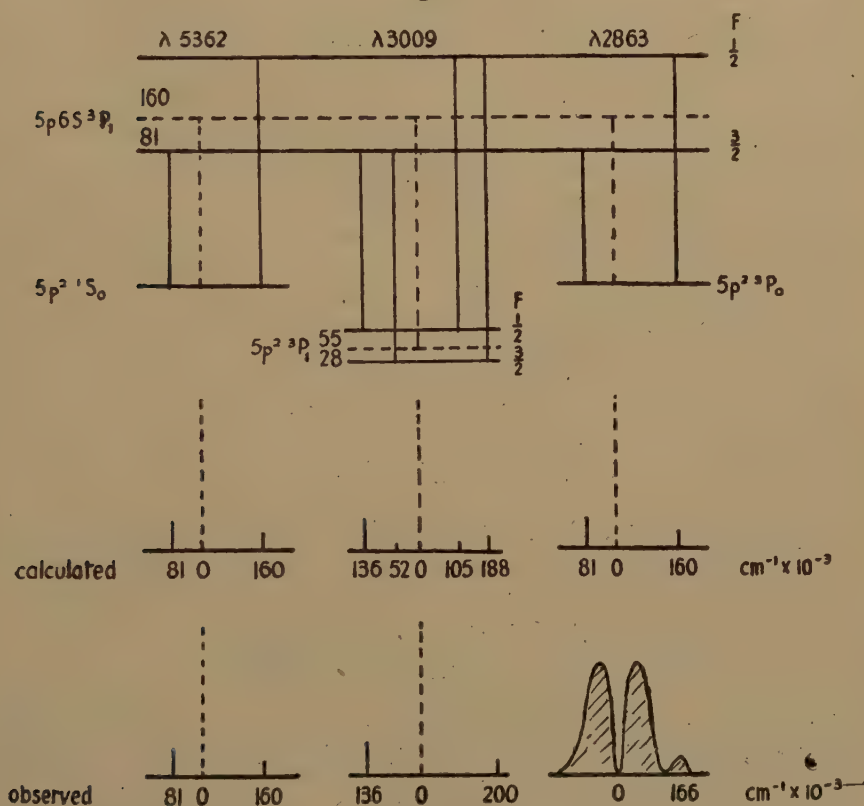
The complete analysis for the lines 5632, 3009, 2863 is shown in fig. 2. The dotted levels and components refer to the even isotopes, and the remaining levels to the odd isotopes. These are treated as having identical spin, $\frac{1}{2}$, and identical magnetic moment. We shall proceed to discuss briefly each line in turn.

5632.—The analysis is highly satisfactory, since the even isotope level falls exactly (within observational error) at the quantum weighed centre of gravity of the odd isotope levels. This structure suffices to prove that the nuclear spins of both the odd isotopes 117, 119, are $\frac{1}{2}$, and that the magnetic moments are identical.

3009.—This line is very weak. The position of the 136 component relative to that of the even isotope line is well established, and has there-

fore been used in evaluating the value of the lower term structure. The two very weak components failed to appear at all on our plates, and the 188 component was itself so weak that it partly merged into the background of impurity spectrum present and could not be measured with great accuracy. Under these conditions we do not therefore consider the difference of 12 units between the observed and predicted positions as a serious contradiction. The actual accuracy in the reported value

Fig. 2.



of the hyperfine structure interval factor for $5p^2^3P_1$ depends upon the 136 component.

2863.—This is a resonance line and, as may be expected under our conditions of excitation, is broad and completely self reversed. One spin component can be observed distant approximately 166 units from the centre of reversal. (No intensities are given since these have no meaning under conditions of reversal.) The agreement between the

predicted and observed patterns is entirely satisfactory, since the 81 component is lost in the broadened even isotope component.

The reversal is extremely sharp, the total width being no more than $37 \text{ cm.}^{-1} \times 10^{-3}$, and possibly less since this is near the limit of resolution of the Lummer plate. There are three strong even isotopes in tin, each contributing to this narrow reversed line. The physical conditions in the state of reversal are not known here, but it is highly improbable that the natural line widths for the individual isotopes can be less than $20 \text{ cm.}^{-1} \times 10^{-3}$. This means that any existing even isotope displacement must be less than $8 \text{ cm.}^{-1} \times 10^{-3}$. It is indeed likely to be much less than this since the observed width may be almost entirely due to natural line width and interferometer width. We are probably safe in putting $5 \text{ cm.}^{-1} \times 10^{-3}$ as an upper limit to any even isotope displacement.

This small value for a line involving the deepest term implies that there is effectively no even isotope displacement in the whole of the spectrum (apart from possible local accidental displacements due to any configuration interaction perturbations). It is clear that *lines* will exhibit no even isotope displacement, if there are almost identical displacements in the upper and lower terms. If all the terms in tin have relatively large isotope displacements then the difference between the deepest term and $5p6s \text{ } ^3P_1$ would almost certainly exceed $5 \text{ cm.}^{-1} \times 10^{-3}$. We therefore conclude that any isotope displacements in the terms of the tin arc spectrum are unobservably small. In this connexion we mention further the line 3034.

3034.—The transition is relatively of high intensity, but is very much broadened and completely reversed. Any hyperfine structure is swamped in the broadening and reversal. The width of the reversed centre is less than $50 \text{ cm.}^{-1} \times 10^{-3}$, and since this line has a greater reversal broadening even than the former, the width of the reversed centre supports the previous conclusions concerning isotope displacement.

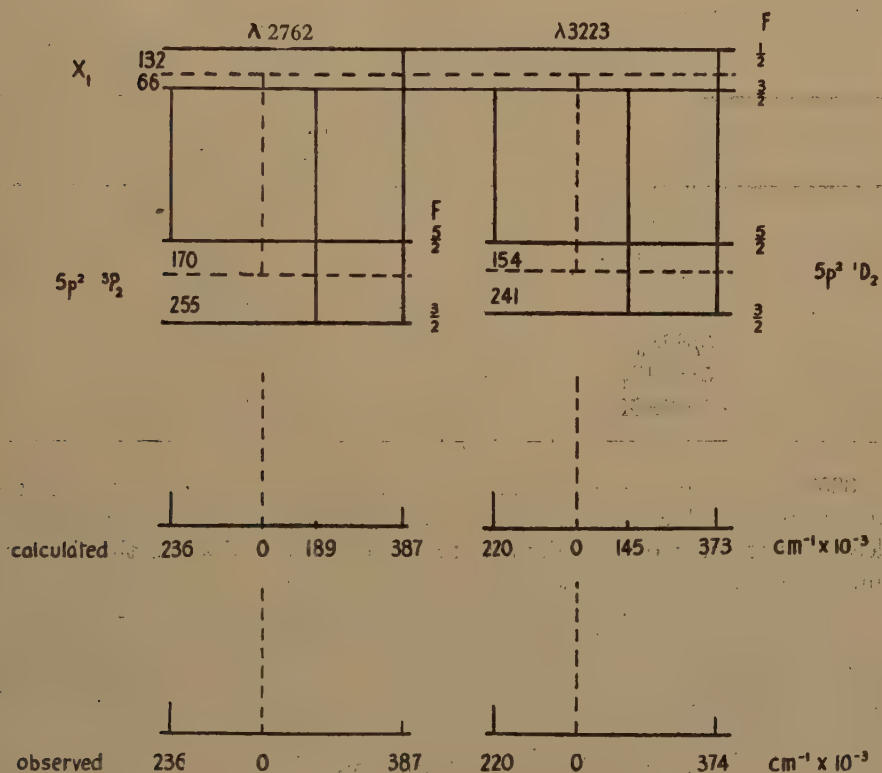
We now come to the discussion of the remaining two classified lines, 3223 and 2762. Unfortunately, the J-value for the common upper term has not yet been given, and the structures for neither of the lower terms are known from the previous lines. Consideration of the intensities and the intervals in 2762 lead us to attribute a J-value of one to the upper term in accordance with the following. Since J is 2 for the lower term, that for the upper can only be 1, 2, or 3. Table III. shows the intensities of the odd isotope components (taking the strongest equal to 10 in each case) for the respective structures when the upper J equals 1, 2, and 3. Neglecting the weaker components, this shows that we have in each case a doublet. The experimental estimated intensity ratio for the doublet in 2762 is 10 : 5. This can only mean that the upper J is equal to 1. Clearly the value J=3 is excluded; yet J=2 might be considered a

possibility. However, analysis shows that if the upper J-value is 2 then the intervals in the structure must have the ratio 3 : 2. For this to hold it would be necessary to displace the weaker component by 31 units, an

TABLE III.

Upper J	Intensities of components.
1	10 : 5.5 : 1.1
2	10 : 6.5 : 0.7 : 0.7
3	10 : 7 : 0.5

Fig. 3.



amount far in excess of the error of observation, which is about 2 units. We therefore conclude that the J-value for the term 19529 cm^{-1} is 1.

The analysis of 2762 and 3223 is shown in fig. 3. **2762.**—The analysis is arrived at in the following manner. A serious

difficulty arises from the fact that only two of the components are observed, and neither of the term structures is known. The observed pattern could therefore be built up out of infinite indeterminate combinations of term structures, either of which may be normal or inverted. Fortunately a restriction is imposed by the fact that the even isotope level in each term is at the quantum weighted centre of gravity, that is, the ratio of the intervals between the even isotope and the F levels is fixed. We can therefore construct a pair of simultaneous equations the solution of which gives a possible structure. This can be done for each of the possible combinations of normal and inverted F-values. When this process is applied, the result leads to those unambiguous single-valued solutions shown in the diagram. Other algebraic solutions give negative values for one or other of the numerical *intervals*, and have therefore no physical meaning.

TABLE IV.

Term.	Total width $\text{cm.}^{-1} \times 10^{-3}$.	Interval factor $\text{cm.}^{-1} \times 10^{-3}$.
$5p^2\ ^3P_0$	0	0
$5p^2\ ^3P_1$	83	55
$5p^2\ ^3P_2$	435	-174
$5p^2\ ^1D_2$	395	-158
$5p^2\ ^1S_0$	0	0
$5p6s\ ^1P_1$	241	161
19529 X_1	198	132

3223.—The measurements made on this line are relatively poor for the line is weak. The intensities are not very reliable. The errors are at least of the order of 10 units and the excellent fit shown in the diagram must be considered fortuitous only. It may be also pointed out that even these badly measured separations are against an upper J-value of 2, for it would be necessary to move the weaker component in by 44 units to accomplish the necessary 3:2 ratio. This is unquestionably very much larger than any possible errors.

Interval Factors.

We give in Table IV. a list of the hyperfine structure interval factors derived from the analysis. In this we have taken into account the fact that the magnetic moment of the nucleus is negative. Hence a structure with inverted F-values has in reality a positive interval factor and is given so in the Table. The negative interval factors arise because of the

directional coupling of the magnetic field of the valency electrons at the nucleus leading to a negative value.

It will be noticed that the $5p^2$ electron group is capable of considerable coupling with the nucleus. This is in interesting agreement with our observations, reported earlier elsewhere, on the high nuclear coupling exhibited by the $5p^4$ electron group in the spectra of iodine.

This investigation has been rendered possible through a grant made to one of us (S. T.) by the Government Grants Committee, for the purchase of instruments, and by a grant made by Messrs. Tootal, Broadhurst, Lee & Co., Ltd. (to G. O. F.).

References.

Tolansky, 'Nature,' cxxxii. p. 318 (1933).

Tolansky, Proc. Roy. Soc. Lond. A, cxliv. p. 574 (1934).

Meggers, Bur. Stand. Journ. Res. xxiv. p. 153 (1940).

XXVI. *Heat Conduction in Composite Circular Cylinders.*

By J. C. JAEGER *.

[Received September 28, 1939.]

THE conduction of heat in composite linear and spherical solids has been extensively studied †. The object of this paper is to present some results for circular cylinders of two materials; it is to be regarded as a sequel to two papers, "A Problem in Conduction of Heat" ‡, and "Some Two-dimensional Problems in Conduction of Heat with Circular Symmetry" §, which will be cited as I. and II., respectively. The method used here is that of the Laplace Transformation as developed in the paper I., which may be consulted for a fuller exposition; the solutions given here are formal, but in all cases it can be verified that the solutions obtained in the form of line integrals satisfy the conditions of Theorem II. of paper I., and the verification that they satisfy the given differential equations and boundary conditions follows as in § 9 of that paper.

The algebra is complicated, but is greatly simplified by the systematic use of cylinder functions; the principal results used are collected in § 2.

In § 3 the hollow composite cylinder with zero initial temperature and boundaries kept at constant temperatures for $t > 0$ is discussed; in § 4 the cylindrical surface source in the hollow cylinder with boundaries kept at zero; in § 5 the corresponding results for the solid composite cylinder are given without proof. A brief discussion of the region outside a circular cylinder is given in § 6.

2. It is convenient to introduce functions defined as follows:—

$$D(x, y) = I_0(x)K_0(y) - K_0(x)I_0(y), \quad (1)$$

$$D_{r,s}(x, y) = \frac{\partial^{r+s}}{\partial x^r \partial y^s} D(x, y), \quad (2)$$

where for brevity $D_1(x, y)$ will be written for $D_{1,0}(x, y)$.

* Communicated by the Author.

† Cf. Carslaw, 'Conduction of Heat' (2nd ed., Macmillan, 1921), §§ 99, 105, 108; Churchill, Duke Math. Journal, ii. p. 405 (1936); *Math. Annalen*, cxv. p. 720 (1938).

‡ Carslaw and Jaeger, Proc. Cambridge Phil. Soc. (in the Press).

§ Carslaw and Jaeger, Proc. London Math. Soc. (in the Press).

These functions are connected with the cylinder functions *

$$C(x, y) = J_0(x)Y_0(y) - Y_0(x)J_0(y), \quad (3)$$

$$C_{r,s}(x, y) = \frac{\partial^{r+s}}{\partial x^r \partial y^s} C(x, y) \quad (4)$$

by the relations

$$D(ix, iy) = -\frac{1}{2}\pi C(x, y), \quad (5)$$

$$i^{r+s} D_{r,s}(ix, iy) = -\frac{1}{2}\pi C_{r,s}(x, y). \quad (6)$$

The following properties are trivial:

$$D_{r+2,s}(x, y) + \frac{1}{x} D_{r+1,s}(x, y) - D_{r,s}(x, y) = 0. \quad (7)$$

$$D_{r,s+2}(x, y) + \frac{1}{y} D_{r,s+1}(x, y) - D_{r,s}(x, y) = 0. \quad (8)$$

$$D(x, y) = -D(y, x). \quad (9)$$

$$D_{0,1}(x, y) = -D_{1,0}(y, x). \quad (10)$$

$$D_{1,0}(x, x) = \frac{1}{x}. \quad (11)$$

$$D(x, y)D_1(x, z) - D(x, z)D_1(x, y) = \frac{1}{x} D(z, y). \quad (12)$$

$$D(x, y)D_{1,1}(x, z) - D_{0,1}(x, z)D_1(x, y) = \frac{1}{x} D_1(z, y). \quad (13)$$

$$D(x, y)D_{1,1}(x, y) - D_{0,1}(x, y)D_1(x, y) = \frac{1}{xy}. \quad (14)$$

$$I_0(x)D_1(x, y) - I_0'(x)D(x, y) = \frac{1}{x} I_0(y). \quad (15)$$

$$I_0(x)D_1(y, z) - I_0(z)D_1(y, x) = I_0'(y)D(x, z). \quad (16)$$

$$I_0(x)D(y, z) + I_0(y)D(z, x) + I_0(z)D(x, y) = 0. \quad (17)$$

Equations (15), (16), (17) are valid also if the I are replaced by K .

3. *Hollow cylinder of one material from $r=a$ to $r=b$ and of another from $r=b$ to $r=c$. The initial temperature of the whole zero. The surface $r=a$ kept at V_0 , and $r=c$ at zero, for $t>0$.*

Let the temperature, conductivity, specific heat, and density in $a<r<b$ be v_1 , K_1 , c_1 , and ρ_1 , and let $\kappa_1 = K_1/c_1\rho_1$; let the corresponding quantities in $b<r<c$ be v_2 , K_2 , c_2 , ρ_2 , and κ_2 .

* It should perhaps be remarked that $C_{23}(\alpha, \beta)$ has been used [Titchmarsh, Proc. Lond. Math. Soc. (2) xxii. p. 15 (1924)] for the function $J_{21}(\alpha)Y_{23}(\beta) - Y_{23}(\alpha)J_{23}(\beta)$, but notations for cylinder functions are not well standardized, and the notation of (3) and (4) is particularly convenient for the type of problem under consideration.

Then we have to solve

$$\kappa_1 \left(\frac{\partial^2 v_1}{\partial r^2} + \frac{1}{r} \frac{\partial v_1}{\partial r} \right) = \frac{\partial v_1}{\partial t}, \quad a < r < b, \quad t > 0, \quad \dots \quad (18)$$

$$\kappa_2 \left(\frac{\partial^2 v_2}{\partial r^2} + \frac{1}{r} \frac{\partial v_2}{\partial r} \right) = \frac{\partial v_2}{\partial t}, \quad b < r < c, \quad t > 0 \quad \dots \quad (19)$$

with $v_1 = v_2, \quad r = b, \quad t > 0. \quad \dots \quad (20)$

$$K_1 \frac{\partial v_1}{\partial r} = K_2 \frac{\partial v_2}{\partial r}, \quad r = b, \quad t > 0. \quad \dots \quad (21)$$

$$v_1 = V_0, \quad r = a, \quad t > 0. \quad \dots \quad (22)$$

$$v_2 = 0, \quad r = c, \quad t > 0. \quad \dots \quad (23)$$

$$v_1 = 0 \text{ when } t = 0 \text{ (} a < r < b \text{)}; \quad v_2 = 0 \text{ when } t = 0 \text{ (} b < r < c \text{)}. \quad (24)$$

Multiplying by e^{-pt} , $p > 0$, integrating with respect to t from 0 to ∞ , and writing

$$\bar{v}_1 = \int_0^\infty e^{-pt} v_1 dt, \quad \bar{v}_2 = \int_0^\infty e^{-pt} v_2 dt,$$

we obtain the subsidiary equations

$$\frac{d^2 \bar{v}_1}{dr^2} + \frac{1}{r} \frac{d\bar{v}_1}{dr} - q_1^2 \bar{v}_1 = 0, \quad a < r < b, \quad \dots \quad (25)$$

$$\frac{d^2 \bar{v}_2}{dr^2} + \frac{1}{r} \frac{d\bar{v}_2}{dr} - q_2^2 \bar{v}_2 = 0, \quad b < r < c, \quad \dots \quad (26)$$

to be solved with

$$\bar{v}_1 = V_0/p, \quad r = a, \quad \dots \quad (27)$$

$$\bar{v}_2 = 0, \quad r = c, \quad \dots \quad (28)$$

$$\bar{v}_1 = \bar{v}_2, \quad r = b, \quad \dots \quad (29)$$

$$K_1 \frac{d\bar{v}_1}{dr} = K_2 \frac{d\bar{v}_2}{dr}, \quad r = b, \quad \dots \quad (30)$$

where $q_1^2 = p/\kappa_1, \quad q_2^2 = p/\kappa_2. \quad \dots \quad (31)$

The general solution of (25) which satisfies (27) is

$$\bar{v}_1 = \frac{V_0}{p} \frac{D(rq_1, bq_1)}{D(aq_1, bq_1)} + A \frac{D(rq_1, aq_1)}{D(aq_1, bq_1)},$$

and the general solution of (26) which satisfies (28) is

$$\bar{v}_2 = BD(rq_2, cq_2).$$

Equations (29) and (30) then require

$$\left. \begin{aligned} AD(bq_1, aq_1) - BD(bq_2, cq_2) &= 0, \\ K_1 q_1 AD_1(bq_1, aq_1) - K_2 q_2 BD_1(bq_2, cq_2) &= -\frac{K_1 V_0}{bp D(aq_1, bq_1)} \end{aligned} \right\}.$$

Solving for A and B and substituting, we obtain finally

$$\bar{v}_1 = \frac{V_0}{p \Delta(p)} \{K_1 q_1 D_1(bq_1, rq_1) D(bq_2, cq_2) - K_2 q_2 D(bq_1, rq_1) D_1(bq_2, cq_2)\}, \quad (32)$$

$$\bar{v}_2 = \frac{K_1 V_0}{bp \Delta(p)} D(rq_2, cq_2), \quad \dots \dots \dots (33)$$

where

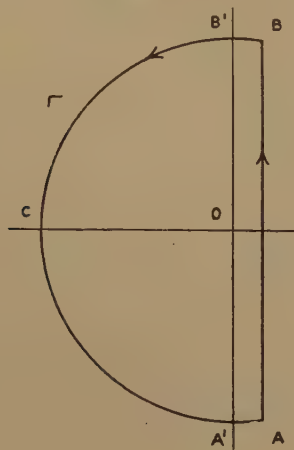
$$\Delta(p) = K_1 q_1 D_1(bq_1, aq_1) D(bq_2, cq_2) - K_2 q_2 D(bq_1, aq_1) D_1(bq_2, cq_2). \quad (34)$$

To evaluate v_2 we have from the inversion theorem for the Laplace Transformation :

$$v_2 = \frac{K_1 V_0}{2\pi i b} \int_{\gamma-i\infty}^{\gamma+i\infty} \frac{e^{\lambda z} D(r\mu_2, \mu_2)}{\lambda \Delta(\lambda)} d\lambda, \quad \dots \dots \dots (35)$$

where $\mu_1 = \sqrt{(\lambda/\kappa_1)}$, and $\mu_2 = \sqrt{(\lambda/\kappa_2)}$. $\dots \dots \dots (36)$

Fig. 1.



The integrand of (35) has simple poles at $\lambda=0$ and $\lambda=-\kappa_1 \alpha_s^2$, where $\pm \alpha_s$, $s=1, 2, \dots$, are the roots of *

$$K_1 C_1(b\alpha, a\alpha) C(\kappa b\alpha, \kappa a\alpha) - \kappa K_2 C(b\alpha, a\alpha) C_1(\kappa b\alpha, \kappa a\alpha) = 0, \quad \dots \quad (37)$$

where $\kappa = \sqrt{(\kappa_1/\kappa_2)}$. $\dots \dots \dots (38)$

Consider the integral of the integrand of (35), taken round the contour of fig. 1, consisting of the line AB distant γ from the imaginary axis and portion of a circle Γ which does not pass through any pole of the integrand. The integral over Γ tends to zero †, as its radius tends to infinity through

* For details of the method of proof see II. § 2.

† C_1 is written throughout for $C_{1,0}$.

a sequence of values which avoid poles of the integrand. Thus the $\int_{\gamma-i\infty}^{\gamma+i\infty}$ in (35) may be replaced by $2\pi i$ times the sum of the residues at the poles of its integrand.

The residue at $\lambda=0$ is

$$\frac{b \log(r/c)}{K_1 \log(b/c) - K_2 \log(b/a)} \quad (39)$$

To evaluate the residue at $\lambda = -\kappa_1 \alpha_s^2$ we have

$$\begin{aligned} 2\lambda \frac{d\Delta(\lambda)}{d\lambda} = & b(K_1 \mu_1^2 - K_2 \mu_2^2) D(b\mu_1, a\mu_1) D(b\mu_2, c\mu_2) \\ & + b\mu_1 \mu_2 (K_1 - K_2) D_1(b\mu_1, a\mu_1) D_1(b\mu_2, c\mu_2) \\ & + c\mu_2 \{K_1 \mu_1 D_{0,1}(b\mu_2, c\mu_2) D_1(b\mu_1, a\mu_1) - K_2 \mu_2 D(b\mu_1, a\mu_1) \\ & \quad \times D_{1,1}(b\mu_2, c\mu_2)\} \\ & + a\mu_1 \{K_1 \mu_1 D(b\mu_2, c\mu_2) D_{1,1}(b\mu_1, a\mu_1) - K_2 \mu_2 D_{0,1}(b\mu_1, a\mu_1) \\ & \quad \times D_1(b\mu_2, c\mu_2)\}. \end{aligned} \quad (40)$$

Now, when $\lambda = -\kappa_1 \alpha_s^2$,

$$\frac{K_2 \mu_2 D_1(b\mu_2, c\mu_2)}{K_1 \mu_1 D_1(b\mu_1, a\mu_1)} = \frac{D(b\mu_2, c\mu_2)}{D(b\mu_1, a\mu_1)} = \frac{C(\kappa b \alpha_s, \kappa c \alpha_s)}{C(b \alpha_s, a \alpha_s)} = \frac{K_2 \kappa C_1(\kappa b \alpha_s, \kappa c \alpha_s)}{K_1 C_1(b \alpha_s, a \alpha_s)} = \sigma, \text{ say,} \quad (41)$$

Introducing this in (40) and using (14), we obtain

$$\begin{aligned} \left[\lambda \frac{d\Delta(\lambda)}{d\lambda} \right]_{\lambda = -\kappa \alpha_s^2} = & \frac{1}{2} \left[b(K_1 \mu_1^2 - K_2 \mu_2^2) \sigma D^2(b\mu_1, a\mu_1) \right. \\ & \left. + b \left(\frac{K_1}{K_2} \right) \mu_1^2 (K_1 - K_2) D_1^2(b\mu_1, a\mu_1) \sigma + \frac{K_1 \sigma}{b} - \frac{K_2}{b\sigma} \right]_{\lambda = -\kappa \alpha_s^2} \\ = & \frac{1}{2bK_2 F(\alpha_s) C(\kappa b \alpha_s, \kappa c \alpha_s) C(b \alpha_s, a \alpha_s)}, \end{aligned} \quad (42)$$

where

$$\begin{aligned} \frac{1}{F(\alpha_s)} = & C^2(\kappa b \alpha_s, \kappa c \alpha_s) \left\{ \frac{1}{4} K_2 b^2 \pi^2 (K_2 \kappa^2 - K_1) \alpha_s^2 C^2(b \alpha_s, a \alpha_s) \right. \\ & \left. + \frac{1}{4} \pi^2 b^2 K_1 (K_1 - K_2) \alpha_s^2 C_1^2(b \alpha_s, a \alpha_s) + K_1 K_2 \right\} - K_2^2 C^2(b \alpha_s, a \alpha_s). \end{aligned} \quad (43)$$

Thus, the residue of the integrand of (35) at $\lambda = -\kappa \alpha_s^2$ is

$$-\pi b K_2 e^{-\kappa_1 \alpha_s^2} C(\kappa r \alpha_s, \kappa c \alpha_s) C(\kappa b \alpha_s, \kappa c \alpha_s) C(b \alpha_s, a \alpha_s) F(\alpha_s). \quad (44)$$

From (39) and (44) we have finally

$$\begin{aligned} v_2 = & V_0 \frac{K_1 \log(r/c)}{K_1 \log(b/c) - K_2 \log(b/a)} \\ & - \pi K_1 K_2 V_0 \sum_{s=1}^{\infty} e^{-\kappa_1 \alpha_s^2} C(\kappa r \alpha_s, \kappa c \alpha_s) C(\kappa b \alpha_s, \kappa c \alpha_s) C(b \alpha_s, a \alpha_s) F(\alpha_s). \end{aligned} \quad (45)$$

In the same way we find from (32), using (12),

$$v_1 = V_0 \frac{K_1 \log(b/c) - K_2 \log(b/r)}{K_1 \log(b/c) - K_2 \log(b/a)} - \pi K_1 K_2 V_0 \sum_{s=1}^{\infty} e^{-\kappa_s \alpha_s t} C(r \alpha_s, a \alpha_s) C^2(\kappa b \alpha_s, \kappa c \alpha_s) F(\alpha_s). \quad (46)$$

For the same problem but with $r=a$ maintained at zero and $r=c$ at V_1 for $t>0$ we find, in place of (32) and (33),

$$\bar{v}_1 = -\frac{K_2 V_1}{bp \Delta(p)} D(r q_1, a q_1),$$

$$\bar{v}_2 = \frac{V_1}{p \Delta(p)} \{K_1 q_1 D(b q_2, r q_2) D_1(b q_1, a q_1) - K_2 q_2 D(b q_1, a q_1) D_1(b q_2, r q_2)\},$$

where $\Delta(p)$ is given by (34). Hence

$$v_1 = -V_1 \frac{K_2 \log(r/a)}{K_1 \log(b/c) - K_2 \log(b/a)} + \pi K_2^2 V_1 \sum_{s=1}^{\infty} e^{-\kappa_s \alpha_s t} C(r \alpha_s, a \alpha_s) C(\kappa b \alpha_s, \kappa c \alpha_s) C(b \alpha_s, a \alpha_s) F(\alpha_s). \quad (47)$$

$$v_2 = V_1 \frac{K_1 \log(b/r) - K_2 \log(b/a)}{K_1 \log(b/c) - K_2 \log(b/a)} + \pi K_2^2 V_1 \sum_{s=1}^{\infty} e^{-\kappa_s \alpha_s t} C(\kappa r \alpha_s, \kappa c \alpha_s) C^2(b \alpha_s, a \alpha_s) F(\alpha_s), \quad (48)$$

where the α_s are the roots of (37).

The solution for the cylinder with $r=a$ maintained at V_0 and $r=c$ at V_1 for $t>0$ is obtained by adding (46) to (47) and (45) to (48).

4. *Hollow cylinder of one material from $r=a$ to $r=b$ and of another from $r=b$ to $r=c$. The surfaces $r=a$ and $r=c$ kept at zero for $t>0$. An instantaneous cylindrical surface source * of strength Q at $t=0$ over $r=r'$, $a<r'<b$.*

Here in the notation of §3 we have to solve (18), (19), (20), (21) with

$$v_1 = 0, \quad \text{when } r=a, \quad t>0, \quad (49)$$

$$v_2 = 0, \quad \text{when } r=c, \quad t>0, \quad (50)$$

$$\text{and } v_1 = u_1 + w_1, \quad (51)$$

$$\text{where } u_1 = \frac{Q}{4\pi\kappa_1 t} \exp[-(r^2 + r'^2)/4\kappa_1 t] I_0\left(\frac{rr'}{2\kappa_1 t}\right), \quad (52)$$

and w_1 satisfies

$$\kappa_1 \left(\frac{\partial^2 w_1}{\partial r^2} + \frac{1}{r} \frac{\partial w_1}{\partial r} \right) = \frac{\partial w_1}{\partial t}, \quad a < r < b, \quad t > 0, \quad (53)$$

$$\text{and } \lim_{t \rightarrow 0} w_1 = 0, \quad a < r < b. \quad (54)$$

* Cf. II., §10.

It is known that *

$$\left. \begin{aligned} \bar{u}_1 &= \frac{Q}{2\pi\kappa_1} I_0(r'q_1) K_0(rq_1), \quad r > r', \\ &= \frac{Q}{2\pi\kappa_1} K_0(r'q_1) I_0(rq_1), \quad r < r' \end{aligned} \right\} \quad (55)$$

The subsidiary equations, formed as in § 3, are

$$\frac{d^2 \bar{v}_2}{dr^2} + \frac{1}{r} \frac{d\bar{v}_2}{dr} - q_2^2 \bar{v}_2 = 0, \quad b < r < c, \quad (56)$$

$$\frac{d^2 \bar{w}_1}{dr^2} + \frac{1}{r} \frac{d\bar{w}_1}{dr} - q_1^2 \bar{w}_1 = 0, \quad a < r < b, \quad (57)$$

$$\bar{v}_2 = 0, \quad r = c, \quad (58)$$

$$\bar{v}_1 = 0, \quad r = a, \quad (59)$$

$$\bar{v}_1 = \bar{v}_2, \quad r = b, \quad (60)$$

$$K_1 \frac{d\bar{w}_1}{dr} = K_2 \frac{d\bar{v}_2}{dr}, \quad r = b, \quad (61)$$

where q_1 and q_2 are defined in (31).

The solution of (56) satisfying (58) is

$$\bar{v}_2 = CD(rq_2, cq_2). \quad (62)$$

As solution of (57) we take

$$\bar{w}_1 = AD(rq_1, bq_1) + BD(rq_1, aq_1). \quad (63)$$

Then, using the value (55) of $\bar{u}_1$, (59) requires

$$AD(aq_1, bq_1) + \frac{Q}{2\pi\kappa_1} I_0(aq_1) K_0(r'q_1) = 0,$$

and (60) and (61) give

$$CD(bq_2, cq_2) - BD(bq_1, aq_1) = \frac{Q}{2\pi\kappa_1} I_0(q_1 r') K_0(bq_1),$$

$$K_2 q_2 CD_1(bq_2, cq_2) - K_1 q_1 BD_1(bq_1, aq_1) = \frac{K_1 q_1 Q}{2\pi\kappa_1} I_0(r'q_1) K_0'(bq_1) + \frac{K_1 A}{b}.$$

Solving for A, B, C and substituting in (62) and (63) we obtain on reduction

$$\bar{v}_2 = \frac{QK_1}{2\pi\kappa_1 b \Delta(p)} D(r'q_1, aq_1) D(rq_2, cq_2), \quad (64)$$

where $\Delta(p)$ is given by (26),

$$\bar{v}_1 = \frac{Q}{2\pi\kappa_1 \Delta(p)} D(rq_1, aq_1) [K_1 q_1 D(bq_2, cq_2) D_1(bq_1, r'q_1) - K_2 q_2 D_1(bq_2, cq_2) D(bq_1, r'q_1)], \quad \text{when } a < r < r' < b, \quad (65)$$

and when $a < r' < r < b$ we interchange r and r' in (65).

* The result follows from Watson, 'Theory of Bessel Functions, § 13.7.

From (64) and the inversion theorem for the Laplace Transformation

$$v_2 = \frac{QK_1}{4\pi^2 i \kappa_1 b} \int_{\gamma-i\infty}^{\gamma+i\infty} e^{\lambda D(r'\mu_1, a\mu_1)D(r\mu_2, c\mu_2)} \frac{d\lambda}{\Delta(\lambda)}, \quad (66)$$

where, as before,

$$\mu_1 = \sqrt{(\lambda/\kappa_1)}, \quad \mu_2 = \sqrt{(\lambda/\kappa_2)}.$$

The integrand of (66) is a single-valued function of λ with simple poles at $\lambda = -\kappa_1 \alpha_s^2$, where the α_s are the roots of (37).

Then, as in § 3, it follows that the line integral in (66) may be replaced by $2\pi i$ times the sum of the residues at these poles. These are immediately evaluated on using (42) and we obtain

$$v_2 = -\frac{1}{4} \pi Q K_1 K_2 \sum_{s=1}^{\infty} \alpha_s^2 e^{-\kappa_1 \alpha_s^2 t} F(\alpha_s) C(b\alpha_s, a\alpha_s) C(r'\alpha_s, a\alpha_s) \\ \times C(\kappa b\alpha_s, \kappa c\alpha_s) C(\kappa r\alpha_s, \kappa c\alpha_s). \quad (67)$$

Similarly, from (65), we obtain, when $a < r < r'$,

$$v_1 = \frac{1}{8} \pi^2 b K_2 Q \sum_{s=1}^{\infty} \alpha_s^3 e^{-\kappa_1 \alpha_s^2 t} F(\alpha_s) C(b\alpha_s, a\alpha_s) C(r\alpha_s, a\alpha_s) C(\kappa b\alpha_s, \kappa c\alpha_s) \\ \times \{K_1 C(\kappa b\alpha_s, \kappa c\alpha_s) C_1(b\alpha_s, r'\alpha_s) - K_2 \kappa C_1(\kappa b\alpha_s, \kappa c\alpha_s) C(b\alpha_s, r'\alpha_s)\}.$$

This becomes, on using (41),

$$v_1 = -\frac{1}{4} \pi K_1 K_2 Q \sum_{s=1}^{\infty} \alpha_s^2 e^{-\kappa_1 \alpha_s^2 t} F(\alpha_s) C(r\alpha_s, a\alpha_s) C(r'\alpha_s, a\alpha_s) C^2(\kappa b\alpha_s, \kappa c\alpha_s), \quad (68)$$

and this, being symmetrical in r and r' , holds for both $a < r < r'$ and $r' < r < b$.

The solution of the problem of an *instantaneous cylindrical surface source of strength* Q at r' , $b < r' < c$, is obtained in the same way.

We find

$$\bar{v}_1 = \frac{QK_2}{2\pi\kappa_2\Delta(p)} D(r'q_2, cq_2) D(rq_1, aq_1), \quad (69)$$

$$\bar{v}_2 = \frac{QD(rq_2, cq_2)}{2\pi\kappa_2\Delta(p)} [K_2 q_2 D(bq_1, aq_1) D_1(bq_2, r'q_2) \\ - K_1 q_1 D_1(bq_1, aq_1) D(bq_2, r'q_2)], \text{ when } c > r > r' > b, \quad (70)$$

where $\Delta(p)$ is given by (34).

Then, from the inversion theorem, we obtain

$$v_1 = -\frac{\pi Q K_2^2 \kappa_1}{4\kappa_2} \sum_{s=1}^{\infty} e^{-\kappa_1 \alpha_s^2 t} \alpha_s^2 F(\alpha_s) C(b\alpha_s, a\alpha_s) C(r\alpha_s, a\alpha_s) \\ \times C(\kappa r'\alpha_s, \kappa c\alpha_s) C(\kappa b\alpha_s, \kappa c\alpha_s), \quad (71)$$

$$v_2 = -\frac{\pi Q K_2^2 \kappa_1}{4\kappa_2} \sum_{s=1}^{\infty} \alpha_s^2 e^{-\kappa_1 \alpha_s^2 t} F(\alpha_s) C^2(b\alpha_s, a\alpha_s) C(\kappa r\alpha_s, \kappa c\alpha_s) \\ \times C(\kappa r'\alpha_s, \kappa c\alpha_s). \quad (72)$$

If we put $Q=2\pi r'f(r')dr'$ in (67), (68), (71), (72) and integrate with respect to r' from a to c , we obtain the solution of the problem in which the initial temperature of the cylinder is $f(r)$ and the surfaces $r=a$ and $r=c$ are kept at zero. Assuming that $f(r)$ is such that the orders of integration and summation may be interchanged, we have

$$v_1 = -\frac{1}{2}\pi^2 K_1 K_2 \sum_{s=1}^{\infty} \alpha_s^2 e^{-\kappa_1 \alpha_s^2 t} F(\alpha_s) C(r\alpha_s, a\alpha_s) C(\kappa b\alpha_s, \kappa c\alpha_s) g(\alpha_s), \quad (73)$$

$$v_2 = -\frac{1}{2}\pi^2 K_1 K_2 \sum_{s=1}^{\infty} \alpha_s^2 e^{-\kappa_1 \alpha_s^2 t} F(\alpha_s) C(b\alpha_s, a\alpha_s) C(\kappa r\alpha_s, \kappa c\alpha_s) g(\alpha_s), \quad (74)$$

where

$$g(\alpha_s) = C(\kappa b\alpha_s, \kappa c\alpha_s) \int_a^b r' f(r') C(r'\alpha_s, a\alpha_s) dr' + \frac{K_2 \kappa_1}{K_1 \kappa_2} C(b\alpha_s, a\alpha_s) \times \int_b^c r' f(r') C(\kappa r'\alpha_s, \kappa c\alpha_s) dr'. \quad (75)$$

5. *The composite solid cylinder of one material, v_1 , K_1 , c_1 , ρ_1 , κ_1 , for $0 \leq r < a$, and of another, v_2 , K_2 , c_2 , ρ_2 , κ_1 , for $a < r < b$.*

The results corresponding to those of §§ 3 and 4 are as follows:—

(i.) *Initial temperature zero. The surface $r=b$ kept at V_0 for $t > 0$.*

$$v_1 = V_0 + \pi K_2^2 V_0 \sum_{s=1}^{\infty} e^{-\kappa_1 \alpha_s^2 t} J_0(r\alpha_s) J_0(a\alpha_s) C(\kappa a\alpha_s, \kappa b\alpha_s) F_1(\alpha_s), \quad (76)$$

$$v_2 = V_0 + \pi K_2^2 V_0 \sum_{s=1}^{\infty} e^{-\kappa_1 \alpha_s^2 t} J_0^2(a\alpha_s) C(\kappa r\alpha_s, \kappa b\alpha_s) F_1(\alpha_s), \quad (77)$$

where $\pm \alpha_s$, $s=1, 2, \dots$, are the roots of

$$K_1 J_0'(a\alpha) C(\kappa a\alpha, \kappa b\alpha) - \kappa K_2 J_0(a\alpha) C_1(\kappa a\alpha, \kappa b\alpha) = 0, \quad (78)$$

where $\kappa = \sqrt{(\kappa_1/\kappa_2)}$, and

$$\frac{1}{F_1}(\alpha_s) = \frac{1}{4}\pi^2 a^2 \alpha_s^2 C^2(\kappa a\alpha_s, \kappa b\alpha_s) [K_2(K_2 \kappa^2 - K_1) J_0^2(a\alpha_s) + K_1(K_1 - K_2) J_0'^2(a\alpha_s)] - K_2^2 J_0^2(a\alpha_s). \quad (79)$$

(ii.) *The surface $r=b$ kept at zero for $t > 0$. Instantaneous cylindrical surface source of strength Q at $t=0$ over $r=r'$, $0 < r' < a$.*

$$v_1 = -\frac{1}{4}\pi Q K_1 K_2 \sum_{s=1}^{\infty} \alpha_s^2 e^{-\kappa_1 \alpha_s^2 t} J_0(r'\alpha_s) J_0(r\alpha_s) C^2(\kappa a\alpha_s, \kappa b\alpha_s) F_1(\alpha_s), \quad (80)$$

$$v_2 = -\frac{1}{4}\pi Q K_1 K_2 \sum_{s=1}^{\infty} \alpha_s^2 e^{-\kappa_1 \alpha_s^2 t} J_0(r'\alpha_s) J_0(a\alpha_s) C(\kappa a\alpha_s, \kappa b\alpha_s) \times C(\kappa r\alpha_s, \kappa b\alpha_s) F_1(\alpha_s), \quad (81)$$

where $F_1(\alpha_s)$ and the α_s are given in (79) and (78).

(iii.) The surface $r=b$ kept at zero for $t>0$. Instantaneous cylindrical surface source of strength Q at $t=0$ over $r=r'$, $a<r'<b$.

$$v_1 = -\frac{\pi Q K_2^2 \kappa_1}{4\kappa_2} \sum_{s=1}^{\infty} \alpha_s^2 e^{-\kappa_s \alpha_s^2 t} J_0(a\alpha_s) J_0(r\alpha_s) C(\kappa r' \alpha_s, \kappa b \alpha_s) \times C(\kappa a \alpha_s, \kappa b \alpha_s) F_1(\alpha_s), \quad (82)$$

$$v_2 = -\frac{\pi Q K_2^2 \kappa_1}{4\kappa_2} \sum_{s=1}^{\infty} \alpha_s^2 e^{-\kappa_s \alpha_s^2 t} J_0^2(a\alpha_s) C(\kappa r \alpha_s, \kappa b \alpha_s) C(\kappa r' \alpha_s, \kappa b \alpha_s) F_1(\alpha_s). \quad (83)$$

(iv.) The surface $r=b$ kept at zero for $t>0$. The initial temperature of the solid $f(r)$.

$$v_1 = -\frac{1}{2} \pi^2 K_1 K_2 \sum_{s=1}^{\infty} \alpha_s^2 e^{-\kappa_s \alpha_s^2 t} J_0(r\alpha_s) C(\kappa a \alpha_s, \kappa b \alpha_s) F_1(\alpha_s) g_1(\alpha_s), \quad (84)$$

$$v_2 = -\frac{1}{2} \pi^2 K_1 K_2 \sum_{s=1}^{\infty} \alpha_s^2 e^{-\kappa_s \alpha_s^2 t} J_0(a\alpha_s) C(\kappa r \alpha_s, \kappa b \alpha_s) F_1(\alpha_s) g_1(\alpha_s), \quad (85)$$

where

$$g_1(\alpha_s) = C(\kappa a \alpha_s, \kappa b \alpha_s) \int_0^a r' f(r') J_0(r' \alpha_s) dr' + \frac{K_2 \kappa_1}{K_1 \kappa_2} J_0(a\alpha_s) \int_a^b r' f(r') C(\kappa r' \alpha_s, \kappa b \alpha_s) dr'. \quad (86)$$

6. The region bounded internally by the cylinder $r=a$. From $r=a$ to $r=b$ the solid is of one material, v_1 , ρ_1 , c_1 , K_1 , κ_1 , and for $r>b$ of another, v_2 , ρ_2 , c_2 , K_2 , κ_2 .

(i.) The surface $r=a$ maintained at temperature V_0 for $t>0$; the initial temperature of the whole zero.

Proceeding as in §3, the Laplace Transforms of the temperatures in $a<r<b$ and $r>b$ respectively are found to be

$$\bar{v}_1 = \frac{V_0}{p\Delta(p)} \{K_1 q_1 K_0(bq_2) D_1(bq_1, rq_1) - K_2 q_2 K_0'(bq_2) D(bq_1, rq_1)\}, \quad (87)$$

$$\bar{v}_2 = \frac{K_1 V_0}{pb\Delta(p)} K_0(rq_2), \quad (88)$$

where

$$\Delta(p) = K_1 q_1 K_0(bq_2) D_1(bq_1, aq_1) - K_2 q_2 K_0'(bq_2) D(bq_1, aq_1), \quad (89)$$

and

$$q_1 = \sqrt{p/\kappa_1}, \quad q_2 = \sqrt{p/\kappa_2}.$$

v_1 and v_2 are determined from these by the use of the inversion theorem. This gives

$$v_2 = \frac{K_1 V_0}{2\pi i b} \int_{\gamma-i\infty}^{\gamma+i\infty} \frac{e^{\lambda t}}{\lambda \Delta(\lambda)} K_0(r\mu_2) d\lambda, \quad (90)$$

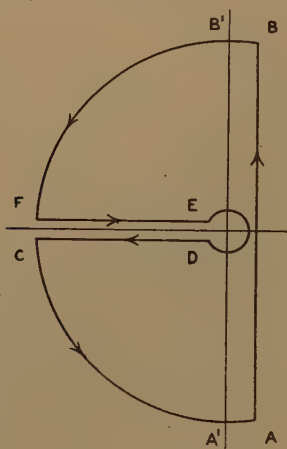
where

$$\mu_1 = \sqrt{(\lambda/\kappa_1)}, \quad \mu_2 = \sqrt{(\lambda/\kappa_2)}.$$

The integrand of (90) has a branch point at the origin, so we use the contour of fig. 2 consisting of the line AB distant γ from the imaginary axis, arcs AA'C and BB'F of a circle Γ whose radius R will tend to ∞ , the lines CD and EF on which $\arg \lambda$ equals $-\pi$ and π respectively, and a small circle about the origin whose radius ϵ will tend to zero. The integral over Γ tends to zero as the radius tends to infinity*. Also there are no poles of the integrand within the contour. Thus, the line integral in (90) may be replaced by the limits of the integrals over CD, EF, and the small circle as $R \rightarrow \infty$ and $\epsilon \rightarrow 0$.

The small circle gives V_0 .

Fig. 2.



On CD we put $\lambda = \kappa_1 u^2 e^{-i\pi}$ and on EF, $\lambda = \kappa_1 u^2 e^{i\pi}$, and obtain finally

$$v_2 = V_0 + \frac{4K_1 V_0}{\pi^2 b} \int_0^\infty \frac{e^{-\kappa_1 u^2} du}{u^2 \{\phi^2(u) + \psi^2(u)\}} [K_1 C_1(bu, au) C(\kappa bu, \kappa ru) - K_2 \kappa C(bu, au) C_1(\kappa bu, \kappa ru)], \quad (91)$$

where

$$\left. \begin{aligned} \phi(u) &= K_1 J_0(\kappa bu) C_1(bu, au) - \kappa K_2 J_0'(\kappa bu) C(bu, au), \\ \psi(u) &= K_1 Y_0(\kappa bu) C_1(bu, au) - \kappa K_2 Y_0'(\kappa bu) C(bu, au), \end{aligned} \right\} \quad (92)$$

and

$$\kappa = \sqrt{(\kappa_1/\kappa_2)}.$$

Similarly, we obtain

$$v_1 = V_0 + \frac{8K_1 K_2 V_0}{\pi^2 b^2} \int_0^\infty \frac{e^{-\kappa_1 u^2} C(ru, au)}{\phi^2(u) + \psi^2(u)} \frac{du}{u^3} \quad (93)$$

* For the method of proof and a more detailed discussion of the procedure of this section, see I.

(ii.) The surface $r=a$ maintained at zero for $t>0$. An instantaneous cylindrical surface source of strength Q at $t=0$ over $r=r'$, $a<r'<b$.

$$v_1 = \frac{2K_1K_2Q}{\pi^3b^2} \int_0^\infty e^{-\kappa_1 u^2} \frac{C(r'u, au)C(ru, au)}{\phi^2(u) + \psi^2(u)} \frac{du}{u}, \quad (94)$$

$$v_2 = \frac{K_1Q}{b\pi^2} \int_0^\infty e^{-\kappa_1 u^2} \frac{C(r'u, au)\chi(r, u)}{\phi^2(u) + \psi^2(u)} du, \quad (95)$$

where

$$\chi(r, u) = K_1C(\kappa bu, \kappa ru)C_1(bu, au) - K_2\kappa C_1(\kappa bu, \kappa ru)C(bu, au), \quad (96)$$

and $\phi(u)$ and $\psi(u)$ are defined in (92).

(iii.) The surface $r=a$ maintained at zero for $t>0$. An instantaneous cylindrical surface source of strength Q at $t=0$ over $r=r'$, $r'>b$.

$$v_1 = \frac{K_2\kappa_1Q}{\pi^2b\kappa_2} \int_0^\infty e^{-\kappa_1 u^2} \frac{C(ru, au)\chi(r', u)}{\phi^2(u) + \psi^2(u)} du, \quad (97)$$

$$v_2 = \frac{\kappa_1Q}{2\pi\kappa_2} \int_0^\infty e^{-\kappa_1 u^2} \frac{\chi(r, u)\chi(r', u)}{\phi^2(u) + \psi^2(u)} u du. \quad (98)$$

(iv.) The surface $r=a$ maintained at zero for $t>0$. The initial temperature of the whole $f(r)$. It is assumed that $f(r)$ is such that the orders of integration may be interchanged.

$$v_1 = \frac{4K_1K_2}{\pi^2b^2} \int_0^\infty e^{-\kappa_1 u^2} \frac{C(ru, au)}{\phi^2(u) + \psi^2(u)} g(u) \frac{du}{u}, \quad (99)$$

$$v_2 = \frac{2K_1}{\pi b} \int_0^\infty e^{-\kappa_1 u^2} \frac{\chi(r, u)}{\phi^2(u) + \psi^2(u)} g(u) du, \quad (100)$$

where

$$g(u) = \int_a^b r'f(r')C(r'u, au)dr' + \frac{\pi b\kappa_1}{2K_1\kappa_2} u \int_b^\infty r'f(r')\chi(r', u)dr'. \quad (101)$$

The University of Tasmania.

XXVII. *Axially Symmetrical Stress in a Thick Tube.*

By A. M. BINNIE, M.A.*

[Received March 6, 1941.]

1. *Introduction.*

THE well-known analysis of the stress in a uniform tube, the wall-thickness of which is comparable to the radius, is confined to cases of plane strain. This paper explains how the stress distribution may be calculated when the restriction is removed, the tube being subjected to an axially symmetrical arrangement of forces. The method employed is an extension of that due to Filon (1902), who examined the stresses in a cylindrical specimen held in a testing machine by various types of grip: it can be applied to a wide variety of cases of imposed stresses and displacements.

The general solution is given in the next section, which is followed by a detailed analysis of the case where the plane ends are free from normal stress and the curved surfaces are subjected to a given distribution of normal and shear stresses. Finally, in section 4, the stresses in a tube loaded with an external belt of uniform pressure are investigated numerically. This example incidentally provides some information concerning the validity of St. Venant's principle.

2. *General Solution.*

With cylindrical co-ordinates r, θ, z (fig. 1), let u and w denote the radial and longitudinal displacements. When u and w are independent of θ , it is known that they are given by:

$$(\lambda + 2\mu) \frac{d^2 u}{dr^2} + (\lambda + 2\mu) \frac{d}{dr} \left(\frac{u}{r} \right) + \mu \frac{d^2 u}{dz^2} + (\lambda + \mu) \frac{d^2 w}{dr \cdot dz} = 0, \quad (2.1)$$

$$(\lambda + \mu) \left(\frac{d^2 u}{dr \cdot dz} + \frac{1}{r} \frac{du}{dz} \right) + \mu \left(\frac{d^2 w}{dr^2} + \frac{1}{r} \frac{dw}{dr} \right) + (\lambda + 2\mu) \frac{d^2 w}{dz^2} = 0, \quad (2.2)$$

where λ and μ are Lamé's elastic constants.

The radial, longitudinal, hoop and shear stresses are respectively:

$$\widehat{rr} = (\lambda + 2\mu) \frac{du}{dr} + \lambda \frac{u}{r} + \lambda \frac{dw}{dz}, \quad (2.3)$$

* Communicated by Professor R. V. Southwell, F.R.S.

$$\widehat{z\bar{z}} = \lambda \frac{du}{dr} + \lambda \frac{u}{r} + (\lambda + 2\mu) \frac{dw}{dz}, \quad . \quad . \quad . \quad . \quad . \quad (2.4)$$

$$\widehat{\theta\theta} = (\lambda + 2\mu) \frac{u}{r} + \lambda \frac{du}{dr} + \lambda \frac{dw}{dz}, \quad . \quad . \quad . \quad . \quad . \quad (2.5)$$

$$\widehat{rz} = \mu \left(\frac{dw}{dr} + \frac{du}{dz} \right). \quad . \quad . \quad . \quad . \quad . \quad (2.6)$$

Filon showed that both du/dz and dw/dr satisfy

$$\left(\frac{d}{dr} \frac{1}{r} \frac{d}{dr} \cdot r + \frac{d}{dz} \right)^2 y = 0, \quad . \quad . \quad . \quad . \quad . \quad (2.7)$$

and that the general solution of (2.7) in terms of product functions is

$$\begin{aligned} y = & A \cos(kz + \alpha) I_1(kr) + B \cos(kz + \beta) K_1(kr) \\ & + C \cos(kz + \gamma) r I_0(kr) + D \cos(kz + \delta) r K_0(kr) \\ & + Ez \cos(kz + \epsilon) I_1(kr) + Fz \cos(kz + \zeta) K_1(kr), \quad . \quad . \quad . \quad . \quad . \quad (2.8) \end{aligned}$$

A, B, etc., α , β , etc., being constants.

We may therefore suppose du/dz and dw/dr to be expressed in series form as :

$$\begin{aligned} \frac{du}{dz} = & \Sigma [A_1 \cos(kz + \alpha_1) I_1(kr) + B_1 \cos(kz + \beta_1) K_1(kr) \\ & + C_1 \cos(kz + \gamma_1) r I_0(kr) + D_1 \cos(kz + \delta_1) r K_0(kr) \\ & + E_1 z \cos(kz + \epsilon_1) I_1(kr) + F_1 z \cos(kz + \zeta_1) K_1(kr)], \quad . \quad (2.9) \end{aligned}$$

$$\begin{aligned} \frac{dw}{dr} = & \Sigma [A_2 \cos(kz + \alpha_2) I_1(kr) + B_2 \cos(kz + \beta_2) K_1(kr) \\ & + C_2 \cos(kz + \gamma_2) r I_0(kr) + D_2 \cos(kz + \delta_2) r K_0(kr) \\ & + E_2 z \cos(kz + \epsilon_2) I_1(kr) + F_2 z \cos(kz + \zeta_2) K_1(kr)]. \quad . \quad (2.10) \end{aligned}$$

3. Solution for a Thick Tube under Imposed Surface Traction.

Consider a thick tube (fig. 1) of length $2c$, external radius a and internal radius b , the applied system of stress being :—

- (i.) $\widehat{rr}/\mu = f(z)$, a given even function of z , over the curved surface $r=a$;
- (ii.) $\widehat{rr}/\mu = F(z)$, a given even function of z , over the curved surface $r=b$;
- (iii.) $\widehat{rz}/\mu = g(z)$, a given odd function of z , over the curved surface $r=a$;
- (iv.) $\widehat{rz}/\mu = G(z)$, a given odd function of z , over the curved surface $r=b$;
- (v.) $\widehat{z\bar{z}}=0$ over the plane ends $z=\pm c$.

Thus u is to be an even function of z and w an odd function of z , and therefore in (2.9) and (2.10)

$$\alpha_1 = \alpha_2 = \beta_1 = \beta_2 = \gamma_1 = \gamma_2 = \delta_1 = \delta_2 = -\pi/2,$$

and

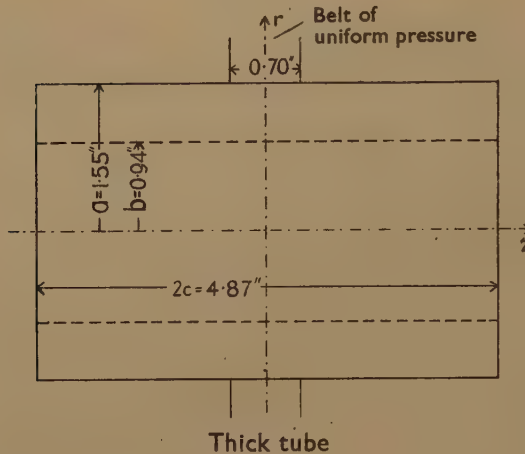
$$\epsilon_1 = \epsilon_2 = \zeta_1 = \zeta_2 = 0.$$

Those equations then become

$$\frac{du}{dz} = \Sigma [A_1 \sin kz I_1(kr) + B_1 \sin kz K_1(kr) + C_1 \sin kz \cdot r I_0(kr) + D_1 \sin kz \cdot r K_0(kr) + E_1 z \cos kz I_1(kr) + F_1 z \cos kz K_1(kr)], \quad (3.1)$$

$$\frac{dw}{dr} = \Sigma [A_2 \sin kz I_1(kr) + B_2 \sin kz K_1(kr) + C_2 \sin kz \cdot r I_0(kr) + D_2 \sin kz \cdot r K_0(kr) + E_2 z \cos kz I_1(kr) + F_2 z \cos kz K_1(kr)], \quad (3.2)$$

Fig. 1.



and on integration we obtain

$$u = \chi(r) + \Sigma \left[-\frac{A_1}{k} \cos kz I_1(kr) - \frac{B_1}{k} \cos kz K_1(kr) - \frac{C_1}{k} \cos kz \cdot r I_0(kr) - \frac{D_1}{k} \cos kz \cdot r K_0(kr) + \frac{E_1}{k} \left(z \sin kz + \frac{\cos kz}{k} \right) I_1(kr) + \frac{F_1}{k} \left(z \sin kz + \frac{\cos kz}{k} \right) K_1(kr) \right], \quad (3.3)$$

$$w = \theta(z) + \Sigma \left[\frac{A_2}{k} \sin kz I_0(kr) - \frac{B_2}{k} \sin kz K_0(kr) + \frac{C_2}{k} \sin kz \cdot r I_1(kr) - \frac{D_2}{k} \sin kz \cdot r K_1(kr) + \frac{E_2}{k} z \cos kz I_0(kr) - \frac{F_2}{k} z \cos kz K_0(kr) \right], \quad (3.4)$$

where $\chi(r)$ and $\theta(z)$ are functions at present unknown.

To determine the relations between the constants we substitute (3.3) and (3.4) in (2.1), and by equating in turn to zero the terms involving $\chi(r)$ and the coefficients of $\cos kz I_1(kr)$, $\cos kz K_1(kr)$, $\cos kz \cdot r I_0(kr)$, $\cos kz \cdot r K_0(kr)$, $z \sin kz I_1(kr)$ and $z \sin kz K_1(kr)$, we obtain

$$\frac{d}{dr} \frac{1}{r} \frac{d}{dr} \{r\chi(r)\} = 0, \quad (3.5)$$

$$(\lambda + \mu)k(A_1 - A_2) + 2(\lambda + 2\mu)C_1 - (\lambda + 3\mu)E_1 - (\lambda + \mu)E_2 = 0, \quad (3.6)$$

$$(\lambda + \mu)k(B_1 - B_2) - 2(\lambda + 2\mu)D_1 - (\lambda + 3\mu)F_1 - (\lambda + \mu)F_2 = 0, \quad (3.7)$$

$$C_1 = C_2 = C \text{ (say)}, \quad (3.8)$$

$$D_1 = D_2 = D \text{ (say)}, \quad (3.9)$$

$$E_1 = E_2 = E \text{ (say)}, \quad (3.10)$$

$$F_1 = F_2 = F \text{ (say)}. \quad (3.11)$$

In a similar manner (2.2) yields

$$\frac{d^2\theta(z)}{dz^2} = 0, \quad (3.12)$$

$$(\lambda + \mu)k(A_1 - A_2) + 2(\lambda + \mu)C_1 + 2\mu C_2 - 2(\lambda + 2\mu)E_2 = 0, \quad (3.13)$$

$$(\lambda + \mu)k(B_1 - B_2) - 2(\lambda + \mu)D_1 - 2\mu D_2 - 2(\lambda + 2\mu)F_2 = 0, \quad (3.14)$$

and four equations identical with (3.8)–(3.11). These last four equations show that (3.6) and (3.13) both reduce to

$$(\lambda + \mu)k(A_1 - A_2) + 2(\lambda + 2\mu)(C - E) = 0, \quad (3.16)$$

and (3.7) and (3.14) to

$$(\lambda + \mu)k(B_1 - B_2) - 2(\lambda + 2\mu)(D + F) = 0. \quad (3.17)$$

From (3.5) we see that

$$\chi(r) = u_0 r + \omega/r, \quad (3.18)$$

where u_0 and ω are constants; and from (3.12), since w is an odd function of z , it appears that

$$\theta(z) = w_0 z, \quad (3.19)$$

where w_0 is a constant.

We may now proceed to satisfy the condition that $\hat{z}\hat{z} = 0$ over the plane ends. On substitution from (3.3), (3.4), (3.8)–(3.11), (3.18) and (3.19), (2.4) becomes

$$\begin{aligned} \hat{z}\hat{z} = & 2\lambda u_0 + (\lambda + 2\mu)w_0 \\ & + \Sigma \left[\left\{ -\lambda A_1 + (\lambda + 2\mu)A_2 - \frac{2\lambda C}{k} + 2(\lambda + \mu)\frac{E}{k} \right\} \cos kz I_0(kr) \right. \\ & + \left\{ \lambda B_1 - (\lambda + 2\mu)B_2 - \frac{2\lambda D}{k} - 2(\lambda + \mu)\frac{F}{k} \right\} \cos kz K_0(kr) \\ & + 2\mu \{ C I_1(kr) - D K_1(kr) \} r \cos kz \\ & \left. + 2\mu \{ -E I_0(kr) + F K_0(kr) \} z \sin kz \right]. \quad (3.20) \end{aligned}$$

For this expression to be zero at $z = \pm c$, we must make

$$u_0 = -(\lambda + 2\mu)w_0/2\lambda, \quad (3.21)$$

$$k = (2n+1)\pi/2c, \text{ where } n=0, 1, 2, \dots, \quad (3.22)$$

$$E = F = 0. \quad (3.23)$$

Hence, from (3.16), (3.17), and (3.23), C and D are given by

$$\frac{2C}{k} = -\frac{(\lambda + \mu)(A_1 - A_2)}{\lambda + 2\mu}, \quad (3.24)$$

$$\frac{2D}{k} = \frac{(\lambda + \mu)(B_1 - B_2)}{\lambda + 2\mu}. \quad (3.25)$$

It is convenient to write

$$\gamma = (\lambda + \mu)/(\lambda + 2\mu). \quad (3.26)$$

After much algebra the expressions (3.3), (3.4), (2.3), (3.20), (2.5) and (2.6) for the displacements and stresses then become :

$$u = u_0 r + \frac{\omega}{r} + \Sigma \frac{\cos kz}{k} \left[A_1 \left\{ -I_1(kr) + \frac{\gamma kr}{2} I_0(kr) \right\} - A_2 \frac{\gamma kr}{2} I_0(kr) \right. \\ \left. - B_1 \left\{ K_1(kr) + \frac{\gamma kr}{2} K_0(kr) \right\} + B_2 \frac{\gamma kr}{2} K_0(kr) \right], \quad (3.27)$$

$$w = w_0 z + \Sigma \frac{\sin kz}{k} \left[-A_1 \frac{\gamma kr}{2} I_1(kr) + A_2 \left\{ I_0(kr) + \frac{\gamma kr}{2} I_1(kr) \right\} \right. \\ \left. - B_1 \frac{\gamma kr}{2} K_1(kr) + B_2 \left\{ -K_0(kr) + \frac{\gamma kr}{2} K_1(kr) \right\} \right], \quad (3.28)$$

$$\frac{\widehat{rr}}{\mu} = -\frac{(3\lambda + 2\mu)}{\lambda} w_0 - \frac{2\omega}{r^2} \\ + \Sigma \cos kz \left[A_1 \left\{ -(1 + \gamma) I_0(kr) + \left(\frac{2}{kr} + \gamma kr \right) I_1(kr) \right\} \right. \\ - A_2 \{ (1 - \gamma) I_0(kr) + \gamma kr I_1(kr) \} \\ + B_1 \left\{ (1 + \gamma) K_0(kr) + \left(\frac{2}{kr} + \gamma kr \right) K_1(kr) \right\} \\ \left. + B_2 \{ (1 - \gamma) K_0(kr) - \gamma kr K_1(kr) \} \right], \quad (3.29)$$

$$\frac{\widehat{zz}}{\mu} = \Sigma \cos kz \left[-A_1 \{ (2\gamma - 1) I_0(kr) + \gamma kr I_1(kr) \} \right. \\ + A_2 \{ (2\gamma + 1) I_0(kr) + \gamma kr I_1(kr) \} \\ + B_1 \{ (2\gamma - 1) K_0(kr) - \gamma kr K_1(kr) \} \\ \left. + B_2 \{ -(2\gamma + 1) K_0(kr) + \gamma kr K_1(kr) \} \right], \quad (3.30)$$

$$\frac{\widehat{\theta\theta}}{\mu} = -\frac{(3\lambda+2\mu)}{\lambda} w_0 + \frac{2\omega}{r^2} + \Sigma \cos kz \left[A_1 \left\{ (1-\gamma) I_0(kr) - \frac{2}{kr} I_1(kr) \right\} - A_2 (1-\gamma) I_0(kr) - B_1 \left\{ (1-\gamma) K_0(kr) + \frac{2}{kr} K_1(kr) \right\} + B_2 (1-\gamma) K_0(kr) \right], \quad (3.31)$$

$$\frac{\widehat{r_z}}{\mu} = \Sigma \sin kz [A_1 \{ -\gamma kr I_0(kr) + I_1(kr) \} + A_2 \{ \gamma kr I_0(kr) + I_1(kr) \} + B_1 \{ \gamma kr K_0(kr) + K_1(kr) \} + B_2 \{ -\gamma kr K_0(kr) + K_1(kr) \}]. \quad (3.32)$$

It will be noticed that, if the terms involving Bessel functions are ignored, the above results reduce to the familiar formulæ for plane strain.

Since both $f(z)$ and $F(z)$ are even functions of z , they can be expanded over the range $z = \pm c$ by Fourier series in the form

$$f(z) - f(c) = \Sigma a_{2n+1} \cos (2n+1) \frac{\pi z}{2c} = \Sigma a_{2n+1} \cos kz, \quad (3.33)$$

$$F(z) - F(c) = \Sigma b_{2n+1} \cos (2n+1) \frac{\pi z}{2c} = \Sigma b_{2n+1} \cos kz, \quad (3.34)$$

where
$$a_{2n+1} = \frac{1}{c} \int_{-c}^c \{f(z) - f(c)\} \cos (2n+1) \frac{\pi z}{2c} dz, \quad (3.35)$$

$$b_{2n+1} = \frac{1}{c} \int_{-c}^c \{F(z) - F(c)\} \cos (2n+1) \frac{\pi z}{2c} dz. \quad (3.36)$$

Similarly, $g(z)$ and $G(z)$, being odd functions of z , may be represented by

$$g(z) = \Sigma s_{2n+1} \sin (2n+1) \frac{\pi z}{2c} = \Sigma s_{2n+1} \sin kz, \quad (3.37)$$

$$G(z) = \Sigma t_{2n+1} \sin (2n+1) \frac{\pi z}{2c} = \Sigma t_{2n+1} \sin kz, \quad (3.38)$$

where
$$s_{2n+1} = \frac{1}{c} \int_{-c}^c g(z) \sin (2n+1) \frac{\pi z}{2c} dz, \quad (3.39)$$

$$t_{2n+1} = \frac{1}{c} \int_{-c}^c G(z) \sin (2n+1) \frac{\pi z}{2c} dz. \quad (3.40)$$

The validity of these expressions in terms of functions of odd multiples of $\pi z/2c$ is discussed in the appendix, for which I am indebted to Professor E. C. Titchmarsh.

Then, on putting $r=a$ in (3.29) and comparing the result with (3.33), we find

$$-\frac{3\lambda+2\mu}{\lambda} w_0 - \frac{2\omega}{a^2} = f(c), \quad (3.41)$$

$$A_1 \left\{ -(1+\gamma) I_0(ka) + \left(\frac{2}{ka} + \gamma ka \right) I_1(ka) \right\} - A_2 \{ (1-\gamma) I_0(ka) + \gamma ka I_1(ka) \}$$

$$+B_1\left\{(1+\gamma)K_0(ka)+\left(\frac{2}{ka}+\gamma ka\right)K_1(ka)\right\} \\ +B_2\{(1-\gamma)K_0(ka)-\gamma kaK_1(ka)\}=a_{2n+1} \quad (3.42)$$

With $r=b$ two similar expressions are obtained from (3.29) and (3.34). Again, from (3.32) and (3.37)

$$A_1\{-\gamma kaI_0(ka)+I_1(ka)\}+A_2\{\gamma kaI_0(ka)+I_1(ka)\} \\ +B_1\{\gamma kaK_0(ka)+K_1(ka)\}+B_2\{-\gamma kaK_0(ka)+K_1(ka)\}=s_{2n+1}, \quad (3.43)$$

with a like result from (3.32) and (3.38). The solution is thus complete, for we have two equations to determine w_0 and ω , and four equations for each set of the coefficients A_1 , A_2 , B_1 , B_2 . No attempt has been made to ensure that $\widehat{rz}=0$ over the plane ends. But this system of radial shears has zero resultant, and St. Venant's principle suggests that its effect is negligible at points not immediately adjoining the ends.

4. Numerical Example.

The purpose of this investigation was to determine the maximum hoop stress on the inner surface of the cylinder shown in fig. 1 when a narrow

TABLE I.

Values of $\frac{a_{2n+1}}{(-p/\mu)}$.

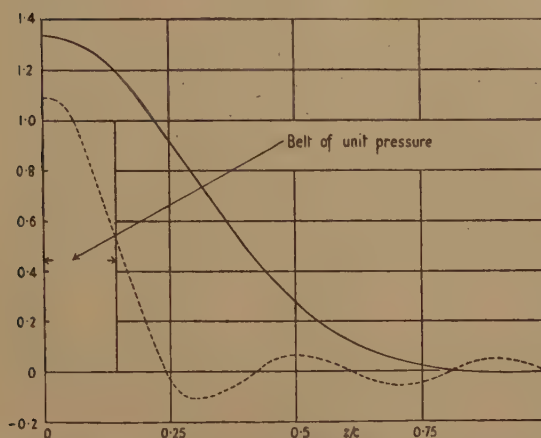
n	0	1	2	3	4
$\frac{a_{2n+1}}{(-p/\mu)}$	0.2851	0.2660	0.2302	0.1819	0.1267

belt of uniform pressure p was applied to the outer surface. The ratio b/a was 0.606, and to shorten the arithmetic the length of cylinder considered was such that $2c=\pi a$. Thus $b_{2n+1}=s_{2n+1}=t_{2n+1}=0$, but the example was a severe test of the feasibility of the method because of the very slow convergence of the Fourier series (3.33). The first five terms of the expansion were, however, found sufficient for the purpose in view.

Tensions are regarded as positive, and the values of $\frac{a_{2n+1}}{(-p/\mu)}$ are given in Table I. the resulting curve being shown by the broken line in fig. 2.

Both $f(c)$ and $F(c)$ were zero, hence w_0 and ω also vanished. Poisson's ratio was taken as $1/4$, so that from (3.26) $\gamma=2/3$. The five sets of simultaneous equations yielded the values of A_1 , A_2 , B_1 , $B_2/-p$ given in Table II.

Fig. 2.



Fourier approximation to belt of unit pressure (---), and ratio of hoop stress over inner surface to applied pressure (—).

TABLE II.

Values of A_1 , A_2 , B_1 , $B_2/-p$.

n	$A_1/-p$	$A_2/-p$	$B_1/-p$	$B_2/-p$
0....	-0.6643	-0.1132	- 0.07502	+ 0.1260
1....	-0.3006	-0.1227	+ 0.3289	+ 3.401
2....	-0.04791	-0.02738	+ 4.370	+12.62
3....	-0.006244	-0.004167	+13.75	+28.49
4....	-0.0007358	-0.0005357	+29.12	+50.94

The coefficients of $\cos (2n+1)\pi z/2c$ for the hoop stress over the inner surface were then found from (3.31) to be :—

TABLE III.

Coefficients of $\cos (2n+1)\pi z/2c$ for $\widehat{\theta\theta}_{r=b}/-p$.

n	0	1	2	3	4
Coefficient of $\cos (2n+1)\frac{\pi z}{2c}$	+0.864	+0.405	+0.075	+0.007	-0.002

These values dwindle much more rapidly than those in Table I., the physical explanation of this result being that (as may be expected from

St. Venant's principle) the effects of pressure variations of small wave-length on the outer surface are insensible at the inner. Thus, one per cent. accuracy was attained with five terms, and the maximum value of $\widehat{\theta\theta}_{r=b}$, which was reached at $z=0$, is found on summing Table III. to be $1.35(-p)$. The distribution of hoop stress over the inner surface is indicated by the continuous curve in fig. 2. If the cylinder length had not extended beyond the loaded portion, the usual simple analysis shows that the hoop stress over the inner surface would have been $3.16(-p)$. Thus the effect of adding an unloaded length was very considerable.

TABLE IV.
Coefficients of $\cos (2n+1)\pi z/2c$ for $\widehat{zz}/-p$ and of
 $\sin (2n+1)\pi z/2c$ for $\widehat{rz}/-p$.

$n \backslash r$	$b=0.606a$	$0.7a$	$0.8a$	$0.9a$	a
Coefficients of $\cos (2n+1)\pi z/2c$ for $\widehat{zz}/-p$.					
0....	-0.117	-0.065	-0.011	+0.044	+0.061
1....	-0.538	-0.281	-0.051	+0.185	+0.474
2....	-0.364	-0.177	-0.045	+0.107	+0.369
3....	-0.167	-0.077	-0.033	+0.036	+0.229
4....	-0.065	-0.030	-0.022	+0.005	+0.141
Coefficients of $\sin (2n+1)\pi z/2c$ for $\widehat{rz}/-p$.					
0....	0	-0.008	-0.010	-0.008	0
1....	0	-0.105	-0.137	-0.103	0
2....	0	-0.113	-0.149	-0.120	0
3....	0	-0.069	-0.056	-0.087	0
4....	0	-0.034	-0.051	-0.056	0

Finally, an attempt was made to determine the approximate magnitudes of the longitudinal and shear stresses, although difficulties due to the discontinuity in the loading were expected. In the region where the load fell abruptly to zero, the stresses of necessity increased to values which theoretically were infinite. For a cylinder Filon (*loc. cit.*) considered in detail the serious mathematical consequences, but for a thick tube it is impracticable to develop algebraical expressions for A_1 , A_2 , B_1 , B_2 , and submit them to rigorous analysis. In practice this point is not of great importance, because a pressure ending sharply cannot in fact be applied. Even if a pressure discontinuity were achieved, the ductility of the material forming the tube would ensure that the local stress did not exceed the yield-point value, which is very far removed from infinity.

The coefficients of $\cos (2n+1)\pi/2c$ for $\widehat{z\bar{z}}/-p$ and of $\sin (2n+1)\pi/2c$ for $\widehat{r\bar{z}}/-p$ were calculated from (3.30) and (3.32), and are shown in Table IV. It will be seen that the series apparently converge, although slowly. Another example of the validity of St. Venant's principle is obtained by comparing the terms forming $\widehat{z\bar{z}}$ at $r=b$ with those at $r=a$. The former, which apply to a region more distant from the load, diminish more rapidly than the latter. The longitudinal stress over the outer surface was compressive and over the inner surface tensile; both were of the same order as the hoop stress considered in Table III. The shear stresses were comparatively small, and with the aid of Table IV, it is possible to show that the shear stress over the plane ends, ignored in the analysis, was actually of negligible magnitude. On multiplying the given figures alternately by $+1$ and -1 , which are the values of $\sin (2n+1)\pi z/2c$ at $z=c$, we notice that the terms nearly cancel each other. Thus in this numerical example no significant error was introduced.

APPENDIX.

Fourier series in terms of odd multiples of $\pi z/2c$ over the range $-\pi < x < \pi$.

The standard form of Fourier series over the range $-\pi < x < \pi$ for an even function $\phi(x)$ is

$$\phi(x) = \frac{A_0}{2} + \sum_{m=1}^{\infty} A_m \cos mx, \quad \dots \dots \dots (i)$$

where

$$A_m = \frac{2}{\pi} \int_0^{\pi} \phi(x) \cos mx \, dx.$$

If, also,

$$\phi(\pi-x) = -\phi(x), \quad \dots \dots \dots (ii)$$

then

$$\begin{aligned} \int_{\pi/2}^{\pi} \phi(x) \cos mx \, dx &= \int_0^{\pi/2} \phi(\pi-y) \cos m(\pi-y) \, dy \\ &= (-1)^{m+1} \int_0^{\pi/2} \phi(y) \cos my \, dy, \end{aligned}$$

and therefore $A_m = 0$ if m is even. Hence (i.) becomes

$$\phi(x) = \sum_{n=0}^{\infty} a'_{2n+1} \cos (2n+1)x, \quad \dots \dots \dots (iii)$$

where

$$a'_{2n+1} = \frac{4}{\pi} \int_0^{\pi/2} \phi(x) \cos (2n+1)x \, dx.$$

Putting $z = \pi z/2c$, and writing $\phi(\pi z/2c) = f(z)$, we obtain

$$f(z) = \sum_{n=0}^{\infty} a'_{2n+1} \cos (2n+1) \frac{\pi z}{2c}, \quad \dots \dots \dots (iv)$$

where
$$a'_{2n+1} = \frac{1}{c} \int_{-c}^c f(z) \cos (2n+1) \frac{\pi z}{2c} dz.$$

This is valid for $-c < z < c$; it holds good in the wider range $-2c < z < 2c$ provided that $f(2c-z) = -f(z)$. The expansion (iv) is valid at the end-point $z=c$ only if $f(c)=0$, since every term on the right-hand side vanishes at $z=c$. Consider then the particular case where $f(z)=f(c)$, ($0 < z < c$); $f(z)=-f(c)$, ($c < z < 2c$). The corresponding series is

$$f(c) = \sum_{n=0}^{\infty} a''_{2n+1} \cos (2n+1) \frac{\pi z}{2c}, \quad \dots \dots \dots \quad (\text{v})$$

where
$$a''_{2n+1} = \frac{1}{c} \int_{-c}^c f(z) \cos (2n+1) \frac{\pi z}{2c} dz.$$

On subtracting (v) from (iv), we obtain

$$f(z) - f(c) = \sum_{n=0}^{\infty} a_{2n+1} \cos (2n+1) \frac{\pi z}{2c}, \quad \dots \dots \dots \quad (\text{vi})$$

where
$$a_{2n+1} = \frac{1}{c} \int_{-c}^c \{f(z) - f(c)\} \cos (2n+1) \frac{\pi z}{2c} dz. \quad \dots \dots \dots \quad (\text{vii})$$

If $f(z)$ is continuous at $z=c$, (vi) holds good at that point because both sides vanish.

Similarly, for an odd function $\psi(x)$ over the range $-\pi < x < \pi$, the standard Fourier series is

$$\psi(x) = \sum_{m=1}^{\infty} S_m \sin mx, \quad \dots \dots \dots \quad (\text{viii})$$

where
$$S_m = \frac{2}{\pi} \int_0^{\pi} \psi(x) \sin mx \, dx.$$

If also
$$\psi(\pi-x) = \psi(x), \quad \dots \dots \dots \quad (\text{ix})$$

then
$$\begin{aligned} \int_{\pi/2}^{\pi} \psi(x) \sin mx \, dx &= \int_0^{\pi/2} \psi(\pi-y) \sin m(\pi-y) \, dy \\ &= (-1)^{m+1} \int_0^{\pi/2} \psi(y) \sin my \, dy, \end{aligned}$$

and therefore $S_m = 0$ if m is even. Hence (viii) becomes

$$(\psi x) = \sum_{n=0}^{\infty} s_{2n+1} \sin (2n+1)x, \quad \dots \dots \dots \quad (\text{x})$$

where
$$s_{2n+1} = \frac{4}{\pi} \int_0^{\pi/2} \psi(x) \sin (2n+1)x \, dx.$$

Putting $x = \pi z/2c$, and writing $\psi(\pi z/2c) = g(z)$, we obtain

$$g(z) = \sum_{n=0}^{\infty} s_{2n+1} \sin (2n+1) \frac{\pi z}{2c}, \quad \dots \dots \dots \quad (\text{xi})$$

where
$$s_{2n+1} = \frac{1}{c} \int_{-c}^c g(z) \sin (2n+1) \frac{\pi z}{2c} dz. \quad \dots \dots (xii)$$

This is valid for the range $-c < x < c$ in any case, and for $-2c < z < 2c$ if $g(2c-z)=g(z)$; it holds good also at $z=c$ if $g(z)$ is continuous there.

Reference.

Filon, L. N. G., Phil. Trans. Roy. Soc. A, cxcviii. p. 147 (1902).

XXVIII. On Algebraic Lamé Functions.

By A. ERDÉLYI, Mathematical Institute of the University, Edinburgh *.

[Received July 22, 1941.]

1. IN a recent note I have proved certain transformation formulæ for Lamé polynomials which in turn implied relations between those characteristic numbers of Lamé's equation which belong to such polynomials. The present note contains the corresponding results for *algebraic* Lamé functions and for the characteristic numbers belonging to such functions. Algebraic Lamé functions are solutions of Lamé's differential equation which are not-rational algebraic functions of $\operatorname{sn} u$, $\operatorname{cn} u$, $\operatorname{dn} u$. Such solutions are known to exist if n is the half of an odd integer and h one of a certain set of $n - \frac{1}{2}$ characteristic values.

2. As in the previous note, I put

$$s = \operatorname{sn}(u, k), \quad c = \operatorname{cn}(u, k), \quad d = \operatorname{dn}(u, k), \quad . \quad . \quad . \quad (1)$$

and use the "Jacobian" form of Lamé's equation,

$$\frac{d^2 y}{du^2} + \{h - n(n+1)k^2 s^2\}y = 0. \quad . \quad . \quad . \quad (2)$$

$2n$ shall be an odd integer ≥ -1 throughout this note. m and p denote not-negative integers. I also use the abbreviations

$$v = K' + i(u - K), \quad \hat{s} = \operatorname{sn}(v, k'), \quad \hat{c} = \operatorname{cn}(v, k'), \quad \hat{d} = \operatorname{dn}(v, k'), \quad . \quad (3)$$

and the relations

$$ks = \hat{d}, \quad ikc = k'\hat{c}, \quad d = k'\hat{s}. \quad . \quad . \quad . \quad (4)$$

For Lamé functions I shall use Ince's notation *as modified in § 8 of my previous note*. According to this, $\operatorname{Ec}_n^{m+\frac{1}{2}}(u, k^2)$ and $\operatorname{Es}_n^{m+\frac{1}{2}}(u, k^2)$ denote mod. $8K$ periodic solutions of (2), having $2m+1$ zeros in the interval $0 \leq u < 4K$, $\operatorname{Ec}_n^{m+\frac{1}{2}}$ being an even and $\operatorname{Es}_n^{m+\frac{1}{2}}$ an odd function of $u-K$. Note that our $\operatorname{Ec}_n^{m+\frac{1}{2}}$ and $\operatorname{Es}_n^{m+\frac{1}{2}}$ are linear combinations of the solutions denoted by the same symbols in Ince's work. This fact is due to shifting the origin from $u=0$ to $u=K$; this change provides a greater symmetry in u and v . The characteristic value of h common

* Communicated by the Author.

to $\text{Ec}_n^{m+\frac{1}{2}}$ and $\text{Es}_n^{m+\frac{1}{2}}$ will be denoted by $a_n^{m+\frac{1}{2}}(k^2)$. As in the previous note, I use the abbreviations

$$\begin{aligned}\text{Ec}_n^{m+\frac{1}{2}} &= \text{Ec}_n^{m+\frac{1}{2}}(u, k^2), & \text{Es}_n^{m+\frac{1}{2}} &= \text{Es}_n^{m+\frac{1}{2}}(u, k^2), & a_n^{m+\frac{1}{2}} &= a_n^{m+\frac{1}{2}}(k^2); \\ \hat{\text{Ec}}_n^{m+\frac{1}{2}} &= \text{Ec}_n^{m+\frac{1}{2}}(v, k'^2), & \hat{\text{Es}}_n^{m+\frac{1}{2}} &= \text{Es}_n^{m+\frac{1}{2}}(v, k'^2), & \hat{a}_n^{m+\frac{1}{2}} &= a_n^{m+\frac{1}{2}}(k'^2).\end{aligned}\quad (5)$$

3. As a preliminary result we shall establish the formulæ

$$(k's+ic)^{\frac{1}{2}}+(k's-ic)^{\frac{1}{2}}=(d+c)^{\frac{1}{2}}+(d-c)^{\frac{1}{2}}, \quad . . . (6)$$

$$(k's+ic)^{\frac{1}{2}}-(k's-ic)^{\frac{1}{2}}=i(d+c)^{\frac{1}{2}}-i(d-c)^{\frac{1}{2}}, \quad . . . (7)$$

$$k's\{(k's+ic)^{\frac{1}{2}}+(k's-ic)^{\frac{1}{2}}\}=d\{(d+c)^{\frac{1}{2}}+(d-c)^{\frac{1}{2}}\}-c\{(d+c)^{\frac{1}{2}}-(d-c)^{\frac{1}{2}}\}, \quad . (8)$$

and

$$ik's\{(k's+ic)^{\frac{1}{2}}-(k's-ic)^{\frac{1}{2}}\}=d\{(d+c)^{\frac{1}{2}}-(d-c)^{\frac{1}{2}}\}-c\{(d+c)^{\frac{1}{2}}+(d-c)^{\frac{1}{2}}\}. \quad (9)$$

All square roots are uniquely determined by the convention that for $u=K$ they shall reduce to the principal value of $k'^{\frac{1}{2}}$.

In order to prove (6), we start with the relation

$$k's+ic+2(k'^2s^2+c^2)^{\frac{1}{2}}+k's-ic=d+c+2(d^2-c^2)^{\frac{1}{2}}+d-c,$$

which is obviously true. Extracting the square-root on both sides and determining the doubtful sign by help of our convention, we obtain (6). The proof of (7) is similar. In order to prove (8) and (9), we multiply the left-hand sides of (6) and (7) by $k's$ and the right-hand sides by $(d^2-c^2)^{\frac{1}{2}}$. Rearranging the terms on the right immediately yields (8) and (9).

4. Disregarding constant factors, the transformation formulae for algebraic Lamé functions read

$$\text{Ec}_n^{m+\frac{1}{2}}=\hat{\text{Ec}}_n^{n-m}, \quad \text{Es}_n^{m+\frac{1}{2}}=\hat{\text{Es}}_n^{n-m}. \quad (10)$$

They imply the relation

$$a_n^{m+\frac{1}{2}}+\hat{a}_n^{n-m}=n(n+1) \quad (11)$$

between the characteristic numbers.

In order to prove (10) and (11), we shall first assume $n=2p+\frac{1}{2}$. From Ince's work it follows that

$$\text{Ec}_{2p+\frac{1}{2}}^{m+\frac{1}{2}}=\{(d+c)^{\frac{1}{2}}+(d-c)^{\frac{1}{2}}\}\Sigma A_r s^{2r}+cd\{(d+c)^{\frac{1}{2}}-(d-c)^{\frac{1}{2}}\}\Sigma B_r s^{2r}, \quad (12)$$

where $\{A_r\}$ and $\{B_r\}$ are two sets of constant coefficients. Using (4), we have

$$\begin{aligned}\text{Ec}_{2p+\frac{1}{2}}^{m+\frac{1}{2}} &= \{(k'\hat{s}+ik'\hat{c}/k)^{\frac{1}{2}}+(k'\hat{s}-ik'\hat{c}/k)^{\frac{1}{2}}\}\Sigma A_r (\hat{d}/k)^{2r} \\ &\quad +i(k'/k)^2\hat{c}k\hat{s}\{(k'\hat{s}+ik'\hat{c}/k)^{\frac{1}{2}}-(k'\hat{s}-ik'\hat{c}/k)^{\frac{1}{2}}\}\Sigma B_r (\hat{d}/k)^{2r},\end{aligned}$$

$$\begin{aligned}\text{or} \quad (k/k')^{\frac{1}{2}}\text{Ec}_{2p+\frac{1}{2}}^{m+\frac{1}{2}} &= \{(k\hat{s}+i\hat{c})^{\frac{1}{2}}+(k\hat{s}-i\hat{c})^{\frac{1}{2}}\}\Sigma A_r (\hat{d}/k)^{2r} \\ &\quad +i(k'/k)^2\hat{c}k\hat{s}\{(k\hat{s}+i\hat{c})^{\frac{1}{2}}-(k\hat{s}-i\hat{c})^{\frac{1}{2}}\}\Sigma B_r (\hat{d}/k)^{2r}.\end{aligned}$$

Using (6) in the first term and (8) in the second term on the right-hand side, we finally have

$$(k/k')^{\frac{1}{2}} \text{Ec}_{2p+\frac{1}{2}}^{m+\frac{1}{2}} = \{(\hat{d}+\hat{c})^{\frac{1}{2}} + (\hat{d}-\hat{c})^{\frac{1}{2}}\} \{\Sigma A_r (\hat{d}/k)^{2r} - (k'/k)^2 \hat{c}^2 \Sigma B_r (\hat{d}/k)^{2r}\} \\ + (k'/k)^2 \hat{c} \hat{d} \{(\hat{d}+\hat{c})^{\frac{1}{2}} - (\hat{d}-\hat{c})^{\frac{1}{2}}\} \Sigma B_r (\hat{d}/k)^{2r}.$$

This last expression is of the form

$$\{(\hat{d}+\hat{c})^{\frac{1}{2}} + (\hat{d}-\hat{c})^{\frac{1}{2}}\} \Sigma \hat{A}_r \hat{s}^{2r} + \hat{c} \hat{d} \{(\hat{d}+\hat{c})^{\frac{1}{2}} - (\hat{d}-\hat{c})^{\frac{1}{2}}\} \Sigma \hat{B}_r \hat{s}^{2r}$$

and thus, apart from a constant factor, represents $\hat{\text{Ec}}_{2p+\frac{1}{2}}^{m'+\frac{1}{2}}$ with some m' as yet unspecified. Determining m' as in section 5 of my previous note, we arrive at the first equation (10) for $n=2p+\frac{1}{2}$.

The corresponding relation for $\text{Es}_n^{m+\frac{1}{2}}$ can be proved similarly, starting with

$$\text{Es}_{2p+\frac{1}{2}}^{m+\frac{1}{2}} = \{(d+c)^{\frac{1}{2}} - (d-c)^{\frac{1}{2}}\} \Sigma A_r s^{2r} + cd \{(d+c)^{\frac{1}{2}} + (d-c)^{\frac{1}{2}}\} \Sigma B_r s^{2r} \quad (13)$$

and using (7) and (9) instead of (6) and (8) respectively,

The case $n=2p-\frac{1}{2}$ is dealt with in a like manner. $\text{Ec}_{2p-\frac{1}{2}}^{m+\frac{1}{2}}$ and $\text{Es}_{2p-\frac{1}{2}}^{m+\frac{1}{2}}$ are represented (*cf.* Ince) by finite series of the form

$$d\{(d+c)^{\frac{1}{2}} \pm (d-c)^{\frac{1}{2}}\} \Sigma A_r s^{2r} + c\{(d+c)^{\frac{1}{2}} \mp (d-c)^{\frac{1}{2}}\} \Sigma B_r s^{2r} \quad (14)$$

From these series the relations (10) can be proved for $n=2p-\frac{1}{2}$.

Incidentally, (10) also shows that all zeros of algebraic Lamé functions are situated on the real axes of u and v and on congruent lines.

(11) follows immediately from (10). $\text{Ec}_n^{m+\frac{1}{2}}$ and $\hat{\text{Ec}}_n^{n-m}$ being constant multiples of each other, their characteristic numbers,

$$a_n^{m+\frac{1}{2}} \quad \text{and} \quad n(n+1) - \hat{a}_n^{n-m}$$

respectively, must be equal. This proves (11).

5. Of some special interest is the case $k^2=\frac{1}{2}$, which is complementary to itself. If $k^2=k'^2=\frac{1}{2}$, (10) reads

$$\text{Ec}_n^{m+\frac{1}{2}}(u) = \text{Ec}_n^{n-m}(v), \quad \text{Es}_n^{m+\frac{1}{2}}(u) = \text{Es}_n^{n-m}(v) \quad (k^2=\frac{1}{2}), \quad (15)$$

while (11) simplifies to

$$a_n^{m+\frac{1}{2}} + a_n^{n-m} = n(n+1) \quad (k^2=\frac{1}{2}). \quad (16)$$

In particular, we have

$$a_{2p+\frac{1}{2}}^{p+\frac{1}{2}} = \frac{1}{2}(2p+\frac{1}{2})(2p+\frac{3}{2}) \quad (k^2=\frac{1}{2}). \quad (17)$$

The last two relations were conjectured by Ince.

References.

- A. Erdélyi, *Phil. Mag.* (7) xxxi. pp. 123-130 (1941).
E. L. Ince, *Proc. Royal Soc. Edinburgh*, lx. pp. 83-99 (1904).

XXIX. *Note on Heine's Integral Representation
of Associated Legendre Functions.*

By A. ERDÉLYI, Mathematical Institute of the University, Edinburgh *.

[Received July 22, 1941.]

1. THE integral of Laplace's type

$$\frac{\Gamma(m+n+1)}{2\pi\Gamma(n+1)} \int_0^{2\pi} \{z+(z^2-1)^{\frac{1}{2}} \cos \phi\}^n \cos m\phi d\phi, \quad \dots \quad (1)$$

representing $P_n^m(z)$ for integral values of m , has been established for positive integer values of n by Heine ('Handbuch der Kugelfunctionen,' Berlin, 1878, vol. i. p. 211): subsequently its validity has been extended to unrestricted values of n (cf. Hobson, 'Spherical and Ellipsoidal Harmonics,' Cambridge, 1931, p. 254). It seems, however, that no corresponding integral representation is known for unrestricted values of m . Hilb (in Pascal's 'Repertorium der höheren Analysis,' Leipzig and Berlin, 1929, vol. I₃, p. 1416) appears to assert that (1) represents $P_n^m(z)$ for unrestricted values of m and $\Re(n) > -1$. As Hilb's statement is not correct, it might be worth while to point out the correct generalization of (1) to arbitrary values of m .

2. We start with Hobson's definition of the associated Legendre function (Hobson, *l. c.* p. 188),

$$P_n^m(z) = 2^{-n-1} \pi^{-1} e^{-\pi i n} \Gamma(-n) \Gamma(m+n+1) (z^2-1)^{\frac{1}{2}m} \\ \times \int_{(C)}^{(z+, 1+, z-, 1-)} (t^2-1)^n (t-z)^{-m-n-1} dt, \quad \dots \quad (2)$$

in which $|\arg(z-1)| < \pi$ is assumed. The contour of integration begins (and ends) at a point C , at which

$$\arg(t-z)=0, \quad |\arg(t-1)| < \frac{1}{2}\pi, \quad |\arg(t+1)| < \frac{1}{2}\pi.$$

Let us suppose that $\Re(z) > 0$ and put

$$t = z + (z^2-1)^{\frac{1}{2}} e^{iu}, \quad \dots \quad (3)$$

where u is uniquely determined by the convention that

$$u=0 \text{ for } t = z + (z^2-1)^{\frac{1}{2}},$$

* Communicated by the Author.

the square-root having its principal value. (2) transforms to

$$P_n^m(z) = (2\pi)^{-2} e^{-\pi i(n+\frac{1}{2})} \Gamma(-n) \Gamma(m+n+1) \\ \times \int_L \{z + (z^2 - 1)^{\frac{1}{2}} \cos u\}^n e^{-m i u} du \quad [\Re(z) > 0] \quad (1)$$

where L is an eight-shaped contour encircling two singularities u_0 and $u_0 - 2\pi$ in the positive and negative direction respectively; u_0 is that value of $-i \log \{-(z-1)^{\frac{1}{2}}(z+1)^{-\frac{1}{2}}\}$, the real part of which lies between 0 and 2π .

3. Replacing n by $-n-1$, we obtain a second integral representation. Integral representations for $e^{im} P_n^m(z)$ can be obtained by replacing u by $u + \omega$ where ω is an arbitrary quantity.

Also we can reduce (4), and similarly the other integrals, to a straight path. Assuming $\Re(n) > -1$ which, since $P_n^m = P_{-n-1}^m$, is no restriction of the generality of our results, we have

$$2\pi \Gamma(n+1) P_n^m(z) = \Gamma(m+n+1) \int_{u_0-2\pi}^{u_0} \{z + (z^2 - 1)^{\frac{1}{2}} \cos u\}^n e^{-m i u} du \\ \times [\Re(n) > -1, \Re(z) > 0]. \quad (5)$$

(4) and (5) are the required generalizations of (1). If m is an integer, (4) or (5) can be reduced to (1) or, in fact, to the integral over any interval of length 2π .

ERRATA.—September 1941.

P. 180, line 26. For $(x-x')^2 + \dots$ read $\{(x-x')^2 + (y-y')^2 + (z-z')^2\}^{\frac{1}{2}}$.

P. 184, line 1. For "then" read "them."

P. 189, line 4 from bottom. For $B_{\mu\nu\tau}^\sigma = 0$ read $B_{\mu\nu\tau}^\sigma = 0$.

P. 196, line 8. For $P(p/q)$ read $P(p|q)$.

Pl. 199, line 10 from bottom. Footnote to $dq dp$. " $\parallel dq dp$ covers all components in the system."

P. 200, line 16. Omit other.

[The Editors do not hold themselves responsible for the views expressed by their correspondents.]

XXX. *The Solubility Relationships of Salts.*

By F. J. GARRICK, M.A.*

[Received July 28, 1941.]

SYSTEMATIC inorganic chemistry provides a number of generalizations concerning the solubilities of salts. These generalizations can be expressed in qualitative or semi-quantitative terms ("sparingly soluble," "readily soluble," "much more soluble than," etc.), and independently of temperature, because of the circumstance that very great differences (millionsfold and greater) are not uncommon, even where there is qualitative similarity of binding forces and where the "molecular" weights are not very different. A comparison of the solubilities of the fluorides and chlorides of potassium and calcium provides a good illustration, especially when it is remembered that such a comparison should apply to the *anhydrous* salts: for the solubility of anhydrous calcium chloride cannot be determined, but must be enormously greater than that of the hydrated salt. The same example shows that there is no *simple* correlation with such factors as valence-type and ionic radius.

This peculiarity of the solubility relationships of salts may be attributed to the circumstance that the solubility of a salt is largely determined by a relatively small difference between two large energy quantities, the lattice energy and the hydration energy of the ions, so that relatively small changes of the variables, if they affect these quantities in different ways, may cause a relatively great change of their difference. This is one of the difficulties of a quantitative theory of solubilities, but offers a point of attack for a theory in such qualitative or comparative terms as have been indicated above, since in comparing the solubilities of two salts it will be possible to disregard all magnitudes whose differences are not large enough to give rise to a large difference of solubility. This means that in a large number of cases only the electrostatic terms need be considered †. Even so, the difficulty remains that it is not as yet possible to express even the electrostatic part of the hydration energy as a function of the charge and radius of an ion; but it is hoped to show in what follows that our knowledge of the character of its dependence

* Communicated by the Author.

† Salts for which this is not true will be called "abnormal" and will be considered separately at a later stage.

on these quantities is sufficient for the deduction of qualitative and comparative solubility relationships.

The Form of the Solubility Function.

The considerations outlined above may be expressed as follows:—In a series of salts, where the individual differences of solubility are not too small, the order of increasing solubility is that of increasing values of a quantity $S = (\text{electrostatic hydration energy of the ions}) - (\text{electrostatic lattice energy})$ *. Here the lattice energy may be written, with sufficient accuracy, $U = Bz^2/(r_+ + r_-)$, where r_+ and r_- are the radii of the ions, z the valency of one of them, and B a constant depending on the crystal structure. The hydration energy of an ion will be denoted by W , with subscripts where necessary to distinguish between ions of different signs or of different valencies, and the hydration energy for all the ions of a salt by ΣW . W is to be regarded as a function of z and r' , where r' is a "corrected" radius of the ion and will be more fully defined later. No attempt will be made to express W in terms of these variables, but its dependence on them will be described qualitatively, by comparison with that of another quantity, $V' \equiv Az^2/r'$.

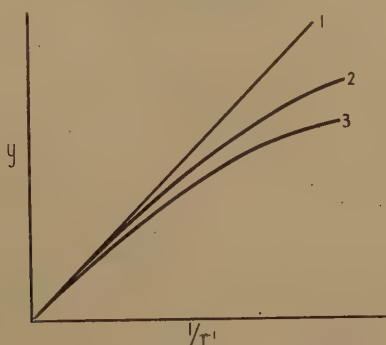
As appears later, r' becomes sensibly equal to r for large values of r , and under these conditions V' becomes equal to $V \equiv Az^2/r$. When A is given the appropriate value, V is simply the well-known "Born charging" formula, which from its derivation must, irrespective of the detailed structure of the medium, give the hydration energy of an "ion" having a very large radius and very small charge. On the other hand, for actual ions in a dipole liquid this expression gives values which are known to be too large, and the same is true of the modified expression V' ; the deviation is of course due to electrical saturation. Accordingly, for large r and small z , $W = V' = V$, but for small r and $z = 1, 2$, etc., $W < V'$. Then, since W is being treated as a continuous function of the variables, *the increase in W resulting from either a decrease of r or an increase of z must be less than the corresponding increase of V'* . This is illustrated (schematically) by fig. 1.

Similar conclusions would result from consideration of *any* model which takes account of the individual water dipoles, for in the immediate neighbourhood of the ion the complete, or at any rate far-reaching, orientation of the dipoles would provide terms in z as well as in z^2 , and involving inverse powers of a distance (*e. g.*, the sum of the radii of an ion and a water molecule) which approaches a finite limit as r approaches zero.

* It is assumed throughout that the solution contains only the same ions as are present in the lattice—*i. e.*, that hydrolysis is negligible.

The use of the "corrected" radius r' depends upon the following considerations. V is independent of the sign of the charge, but there is evidence that this is not true for W . Not only various estimates of "empirical" hydration energies, but also other properties (*e. g.*, ionic mobilities), suggest that the closest correspondence obtains not between cations and anions of equal radii but between pairs such that the anion is somewhat the larger. In terms of a model which represents the water molecule as a spherical dipole carrier this can be provided for by noting that the dipole need not necessarily coincide with the centre of the sphere; and in fact the "centre of the sphere" means the centre of the electron charge cloud, and so is to be identified not with the mid-point, but with the negative end, of the dipole axis.

Fig. 1 (schematic).



Curve 1. $y = V'/z^2$, ($z=1, 2$, etc.).

Curve 2. $y = W/z^2$, ($z=1$).

Curve 3. $y = W/z^2$, ($z=2$).

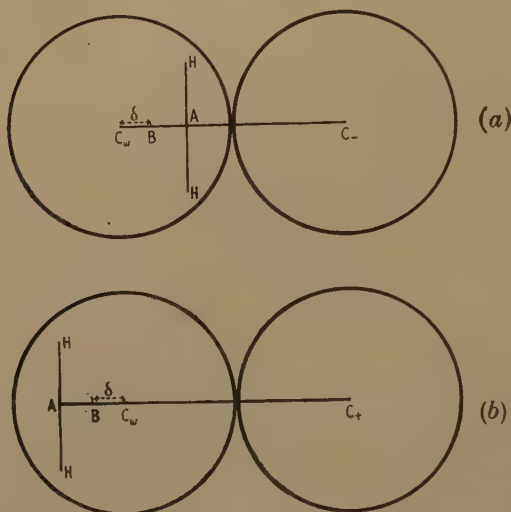
Accordingly the association of a water molecule with an anion and a cation respectively may be represented as in fig. 2 (a) and (b).

Here C_- , C_+ and C_ω are the centres of the ions and of the water molecule respectively, and of course $C_-C_\omega = r_- + r_\omega$, $C_+C_\omega = r_+ + r_\omega$. HH are the hydrogen nuclei, and A , the mid-point of HH , is the positive end of the dipole, whilst C_ω is its negative end, and B is the mid-point of AC_ω , BC_ω being denoted by δ . The ion-dipole distances are $C_-B = r_- + r_\omega - \delta$ and $C_+B = r_+ + r_\omega + \delta$, which may be written $r_\omega + r'$, where $r'_- = r_- - \delta$ and $r'_+ = r_+ + \delta$. The hydration energy expressed as a function of r' , $W(r')$, is now independent of the sign of the charge, and its dependence upon z and r' may be described as above with reference to the modified quantity $V' = Az^2/r'$, and as illustrated in fig. 1. It is obvious but important that $r'_+ + r'_- = r_+ + r_-$, and it may also be noted that when r is very large, so that $W = V$, then $W(r) = W(r')$ and $V = V'$. From known constants of

the water molecule δ is 0.2–0.3 Å.* In what follows it will be taken as 0.25 Å. This choice is not at all critical: any value between 0.2 and 0.3 Å. would lead to the same conclusions.

To summarize: the order of the solubilities of salts, so far as the larger differences are concerned, will be that of the quantities $S = \Sigma W - U$, where $U = Bz^2/(r_+ + r_-)$, W is a function of z and r' whose dependence on these variables may be qualitatively described in terms of that of $V' = Az^2/r'$ as illustrated in fig. 1, and where $r'_- = r_- - 0.25$ Å., and $r'_+ = r_+ + 0.25$ Å. From this it is possible to deduce qualitatively the effects on S of certain changes of the variables, and to compare these

Fig. 2.



effects with observed solubility relationships. For this purpose it is to be noted that since only large differences of solubility, *i. e.*, relatively large differences of S , are under consideration, and since S is always small compared to ΣW and U , δS may be taken as having the same sign as $\delta \Sigma W / \Sigma W - \delta U / U$, and may be neglected when this is zero.

The Effects on S of Certain Changes of the Variables.

The Valency Effect.—Let the radii of the ions, their relative numbers, and the crystal structure, all remain constant, while the valency of each ion increases in a fixed ratio. (In other words, let the valence-type

* Cf., *e. g.*, Bernal and Fowler. It is of interest to compare their treatment with this.

change from 1 : 1 to 2 : 2 or from 1 : 2 to 2 : 4, etc., while other factors remain constant.) Then U increases proportionally to z^2 , and so does V' ; but as has been shown, the increase of W is less than that of V' , and therefore less than that of U ; consequently S decreases. *That is, increasing the valencies of the ions in the same ratio, ceteris paribus, leads to a decrease of solubility.*

The Radius-sum Effect.—Let the valencies, and the ratio ($\alpha = r'_-/r'_+$) of the "corrected" radii, remain constant, while the radii of both ions decrease. Then $\Sigma V'$ takes the form $(1 + 1/\alpha)Az^2/r'_+$ and the denominator of U becomes $(r'_+ - \delta) + (\alpha r'_+ + \delta) = r'_+(1 + \alpha)$, so that $\Sigma V'$ and U are both inversely proportional to r' . Now, on decreasing r' , the increase of ΣW is less than that of $\Sigma V'$, and therefore less than that of U , so that S decreases. *That is, decreasing the radii of the ions leads, ceteris paribus, to a decrease of solubility.*

The Radius-Ratio Effect for Salts of Symmetrical Valence Type.—Let the radii of the ions vary while their sum remains constant. Then, if the valencies are fixed, and the crystal-structure does not change, U is constant, and ΣW has the form $f(r') + \phi(\text{const.} - r')$. For the symmetrical valence-type the two ions are of equal charge and present in equal numbers, so that $f \equiv \phi$. Then there must be a turning-point (minimum or maximum) of ΣW , and accordingly of S , when $r'_+ = r'_-$. It is reasonable to assume, though difficult to prove, that the functions W and V' are not so very different that their second derivatives with respect to r' should be of opposite sign; then, since $\partial^2 V' / (\partial r')^2$ is positive, $\partial^2 W / (\partial r')^2$ will also be positive, and the turning-point will in fact be a minimum. *That is, for salts of symmetrical valence-type the solubility will be a minimum, ceteris paribus, when the "corrected" radii of the ions are equal.*

The Radius-Ratio Effect for Salts of Unsymmetrical Valence Types.—Considerations analogous to those of the preceding paragraph show that for any valence type there will be some value of the radius-ratio for which the solubility will be a minimum. In order definitely to locate the minimum a knowledge of the dependence of W on z is required, since we are dealing with two ions of different valency. The most interesting of the unsymmetrical types is the 1 : 2 type, and this alone will be considered. Using subscripts 1 and 2 to denote quantities referring to the uni- and bivalent ions respectively, and writing $r_2' + r_1' = a$, we have $\Sigma W = W_2(r_2') + 2W_1(a - r_2')$, and the position of the minimum will be defined by $\partial W_2 / \partial r_2' = 2\partial W_1 / \partial r_1'$. Now, if it were true for W (as it is for V') that $W_2 = 4W_1$, this would define the minimum as being where $r_2' = \sqrt{2} \cdot r_1'$. But, as shown above $W_2 < 4W_1$; on the other hand, $W_2 > 2W_1$ (this follows, e. g., from consideration of a dipole model, with some terms in z

and some in z^2), and with only this information it is only possible to say that the minimum will occur where $r_1' < r_2' < \sqrt{2} \cdot r_1'$. For the sake of definiteness we may take the mean of the two limits, and say that *in salts of valence-type 1 : 2 there will, ceteris paribus, be a minimum of solubility where the "corrected" radius of the bivalent ion is greater than that of the univalent ion by a factor in the region of 1.2.* A further point is that the curve of S as a function of the radius of one of the ions (for constant radius-sum) is not symmetrical about the minimum, the branch for small radii of the bivalent ion being much steeper than that for small radii of the univalent ion.

Comparison with Observation.

The considerations which have been brought forward refer to the solubilities of anhydrous salts, and in what follows all numerical data refer to these. Where no other indication is given, experimentally determined values (taken from Landolt-Bornstein) of the solubility of the anhydrous salt at room temperature are quoted. The figures are moles per litre, and are rounded off as high exactitude is superfluous for this purpose. In some other instances estimates of the solubility of the anhydrous salt have been obtained by extrapolation from higher temperatures at which it is the stable phase: these are indicated by the sign $\sim$. In a few other instances figures for the solubility of the lowest hydrate, either as determined at, or as extrapolated to, room temperature, are given with the sign $\gg$, since the solubility of the anhydrous salt must be considerably greater.

The Alkali Halides.—Table I. shows the solubilities of the alkali halides, together with the radii and the "corrected" radii of their ions. The table is so arranged that the diagonal from bottom right-hand to top left-hand is the locus of nearly equal "corrected" radii, and is the direction of decreasing radius sum. The occurrence of low solubilities along this diagonal, and of high solubilities near the other two corners, corresponds exactly with the prediction of the radius-ratio effect. Further, if CsBr and CsI are excluded, there is a clear trend of decreasing solubility along this diagonal in the direction of decreasing radii, as required by the radius-sum effect. The outstandingly low solubility of LiF corresponds with the circumstance that it combines the lowest radius-sum with nearly equal "corrected" radii, and the high solubility of KF, with ions whose actual radii are very nearly equal, emphasizes the importance of using the "corrected" radii. The caesium salts with large anions have solubilities rather lower than accords with the general trend. This is connected with the appearance of the "CaCl" structure, and is to be attributed to the presence in U of significant

van der Waals terms, so that to this extent these salts may be considered to show "abnormal" solubility. The point will be referred to again in a later section.

TABLE I.

Cation ↓	Anion →		F'		Cl'	Br'	I'
	$r \rightarrow$	$r' \rightarrow$					
	$r \downarrow$	$r' \downarrow$	$r' \rightarrow$				
Li ⁺	0.78	1.03	0.07		~24	≥ 24	≥ 26
Na ⁺	0.98	1.23	1.0		5.6	~11	~18
K ⁺	1.33	1.58	21		4.5	5.3	9
Rb ⁺	1.49	1.74	≥ 29		7.1	6.0	6.8
Cs ⁺	1.65	1.90	≥ 38		11	5.1	2.7

$r' + \sim r$

The Alkaline-Earth Halides.—The radii, "corrected" radii, and solubilities are shown in Table II. The beryllium salts are not included, since they are very largely hydrolyzed in solution, and in fact no quantitative data are available. Qualitatively, they are all extremely soluble, in accordance with their extreme radius ratios. With regard to the other salts, we expect to find minimal solubility

where the bivalent ion (here the cation) has a "corrected" radius approximately 1.2 times greater than that of the univalent ion (here the anion). This is very nearly true for one salt only, CaF_2 , and this is the least soluble of the group. Then, for radii of the bivalent ion larger than correspond to this value, we expect to find slowly increasing solubilities, and this is fulfilled by the only salts to which it applies, namely SrF_2 and BaF_2 . Finally, when the radius of the bivalent ion falls considerably below the value corresponding to the minimum, we are dealing with the steeper branch of the curve and expect to find the solubilities rapidly increasing. This is the case with MgF_2 , and with all of the chlorides, bromides, and iodides. These latter all exist as very stable hydrates, which are readily soluble, so that the solubilities

TABLE II.

			Anion	F.	Cl.	Br.	I.
			r	1.33	1.81	1.96	2.20
			r'	1.08	1.56	1.71	1.95
Cation.	r .	r' .					
Mg	0.68	0.93	2×10^{-3}	very soluble, stable hydrates.			
Ca	1.05	1.30	2×10^{-4}	"	"	"	"
Sr	1.16	1.41	3×10^{-3}	"	"	"	"
Ba	1.31	1.56	7×10^{-3}	"	"	"	"

of the anhydrous salts must be very high, though it is impossible to obtain even the roughest quantitative estimate of them. Consideration of the degree of hydration, and of the stabilities and solubilities of the hydrates, would suggest that those having the smaller cations and the larger anions have the highest solubility, as predicted. MgF_2 , while considerably more soluble than CaF_2 , is much less soluble than the rest of this group, and this accords with its small radius-sum, as in the case of lithium fluoride.

Oxides and Sulphides.—For the oxides and sulphides of the alkaline earth metals, the valence-effect would lead to the expectation of solubilities considerably lower than those of the fluorides and chlorides of the corresponding alkali metals, while those of the alkali metals would be expected to have solubilities comparable with those of alkaline-

earth chlorides or bromides (for here the bivalent ion is an anion, so that its "corrected" radius is *less* than its actual radius). Unfortunately, the oxide and sulphide ions are so largely hydrolyzed in solution that there are no data by which to verify these expectations. The slowness of the reaction of magnesium oxide with water would seem to imply that its "solubility" must be very low, as would be expected from comparison in the light of the valence-effect with lithium fluoride, of which the ions have similar radii. The very low solubilities of the oxides and sulphides of the heavier metals are of course "abnormal" solubilities.

Salts of Oxyacids: Perchlorates and Sulphates.—The considerations which have been brought forward apply strictly only to spherical, *i. e.*, monatomic ions, but an extension to polyatomic ions that have an approximation to spherical symmetry is of interest, if due caution be observed. It would not be justifiable to compare such ions with monatomic ions, because of differences in respect of thermal magnitudes, as well as of modes of packing in a crystal or in solution; and it is not possible to assign definitely an effective radius to such an ion. But it would seem possible to compare among themselves the different salts of one such ion, and further to consider two such ions in which different central atoms are united with the same number of the same kind of outer atoms as being similar in "size" and shape, and so to compare their salts. The perchlorate and sulphate ions are such a pair, and may, as an approximation, be considered to differ only in charge. Accordingly, data on the solubilities of the perchlorates and sulphates of the alkalis and the alkaline earths are given in Table III. First, in the light of the *valence effect*, we may compare the alkali perchlorates with the alkaline-earth sulphates, choosing pairs of salts in which the cations have similar radii, and should expect to find the alkaline-earth sulphate always much less soluble than the alkali perchlorate; this is fulfilled wherever comparison is possible: as with NaClO_4 and CaSO_4 , or with KClO_4 and BaSO_4 . Second, the *radius-ratio effect for salts of symmetrical valence-type* leads to the expectation, in both groups (M^+ClO_4 and M^{II}SO_4), of minimal solubility where the actual radius of the cation is less than that of the anion by 0.5 Å. (equal "corrected" radii). Without postulating a definite value for its radius, we must regard the anion as large, so the minimum would only occur for a large cation. In fact, we find in the alkaline earth sulphates a steady decrease of solubility with increasing radius of the cation (the largest cation being Ba^{++} , $r = 1.31$), while in the alkali perchlorates we find, with increasing radius of the cation, solubilities decreasing to a minimum for RbClO_4 ($r_+ = 1.49$). The relationships in these two groups are thus in complete agreement with

expectation, and further, would be consistent with an effective radius for the SO_4^{--} or the ClO_4^- ion of approximately 2 Å. Finally, the *radius-ratio effect for salts of valence-type 1:2* leads to the expectation in the other two groups (alkali sulphates and alkaline-earth perchlorates) of minimal solubility where the "corrected" radius of the bivalent ion is greater than that of the univalent ion by a factor in the neighbourhood of 1.2. In the alkaline-earth perchlorate group, where the univalent ion is the polyatomic anion, this condition could only be fulfilled for a very large cation indeed: putting tentatively

$$r_+' = r_+ + 0.25 = 1.2r'_- = 1.2(r_- - 0.25),$$

and inserting the effective radius of approximately 2 Å. indicated above for the anion, we find $r_+ = 1.85$ Å. That is, the alkaline-earth perchlorates are salts in which the bivalent ion is much smaller than corresponds with the minimum, so that their very high solubilities are in full accordance with expectation. On the other hand, in the alkali-sulphate group the condition for the minimum is that the polyatomic anion should

TABLE III.

Cation. <i>r</i>	Li. 0.78	Na. 0.98	K. 1.33	Rb. 1.49	Cs. 1.65	Mg. 0.68	Ca. 1.05	Sr. 1.16	Ba. 1.31
ClO_4	s. h.*	—	0.1	0.05	0.1	very soluble stable hydrates †.			
SO_4	„	3.6	0.6	1.4	5	s. h.	0.01	10^{-3}	10^{-5}

have somewhat the larger "corrected" radius, and this might be realized for a moderately large cation. In fact, in this group, we find with increasing radius of the cation solubilities decreasing to a minimum for K_2SO_4 and then increasing again. Inserting the radius of the potassium ion (1.33 Å.) in the equations

$$r'_- = (r_- - 0.25) = 1.2r_+' = 1.2(r_+ + 0.25),$$

we find as a tentative estimate of the effective radius of the polyatomic anion $r_- = 2.15$ Å., which accords very well with that obtained above. The observed relationships in all these groups are thus in satisfactory agreement with the theory.

* s. h. = very soluble, stable hydrate.

† Landolt-Bornstein records values for the solubilities of these salts with the anhydrous salt as solid phase. Reference to the original paper shows this to be erroneous (Willard and Smith, Journ. Amer. Chem. Soc. cxxlv. p. 86 (1923), and indeed the Mg and Ba salts are used as dehydrating agents, comparable to P_2O_5 !

Salts of other oxyacids for the most part fall outside the range of applicability of the considerations of this paper, either because of extensive hydrolysis, as with the tertiary orthophosphates, or because of wide departure from a spherical configuration, as with nitrates.

Complex Salts.—Certain complex ions may be considered as approximating to spherical symmetry, and the solubility relationships of their salts may be discussed with the same limitations as in the case of the four-co-ordinated oxyanions. The hexacido anions, and the hexammino cations, are examples, and of course they are so very much larger than any monatomic cation that the condition for minimal solubility will never be fulfilled. Accordingly we expect to find, among the halides of any hexammino cation, and among the alkali or alkaline-earth salts of any hexacido anion, a decrease of solubility with increasing radius of the monatomic ion. Data are scanty, but so far as they are available are in complete accordance with this expectation. Groups in which the data are complete include the alkali fluosilicates and chloroplatinates, and the hexamminocobaltic halides (*cf.* Landolt-Bornstein). Further, the valence effect would lead to the expectation of a decrease in solubility in passing from, *e.g.*, alkali chloro-antimonates to alkaline-earth chloroplatinates, or from hexammino-nickel perchlorate to hexamminoplatinic sulphate. In all such cases the qualitative descriptions of the solubility (quantitative data being usually not available) are consistent with expectation.

Abnormal Solubility.—In many cases very great differences of solubility are observed between salts of cations having electron-shells of other than inert-gas types, and the corresponding salts of the alkali or alkaline-earth cations of the same charge and similar radii. The most striking examples are the very low solubilities of silver, cuprous, and thallic halides (except the fluorides), lead halides, and the oxides and sulphides of these and other heavy metals. This low solubility has been generally attributed to the appearance of relatively large van der Waals terms in the lattice energy, and the order of decreasing solubility in all of these cases is, consistently with this view, that of increasing polarizability of the anion, *viz.*, Cl, Br, I, S. On the other hand, the fluorides, nitrates, and perchlorates of the non-inert-gas cations are often much *more* soluble than the comparable alkali or alkaline-earth salt: compare silver fluoride, nitrate or perchlorate with the corresponding sodium salts, and zinc fluoride with magnesium fluoride. The polarizability of the fluoride ion is less than that of a water molecule, so that possibly van der Waals terms in solution, greater than those in the lattice, should be allowed for; but this is hardly sufficient to account for the whole of the rather complicated relationships. Further, the solubilities of many salts of

the non-inert-gas cations are not abnormal ; there is a close resemblance in solubility between the sulphates and the nitrates of zinc and magnesium or of lead and barium.

Conclusions.

It would seem that the theory which has been put forward gives an adequate account of the solubility relationships of salts of ions of inert-gas type, and of those polyatomic and complex ions which show some approximation to spherical symmetry ; but that with cations of other than inert-gas type specific effects sometimes occur, which cannot be completely accounted for at present.

Summary.

Consideration of the ways in which the hydration energies of ions and the lattice energies of salts are affected by variation of the ionic charges and radii leads to the deduction of a number of generalizations about the relative solubilities of salts, and these are compared with observation.

Inorganic Chemistry Department,
The University, Leeds.

XXXI. *A Derivation of the Equations of Equilibrium of a Thin Plate.*

By HAROLD JEFFREYS *.

[Received December 13, 1940.]

THE treatment of thin plates in the standard books on elasticity is very complicated. It can be much shortened by using the principle of virtual work (also used in Rayleigh's 'Sound') and tensor notation. If w is the displacement normal to the undisturbed position of the central plane, and W the elastic energy per unit surface, w and W are independent of the directions of the axes of x and y , and therefore are scalar functions of the two-dimensional tensor x_i ($i=1, 2$) representing position in the plane. Then $\partial^2 F / \partial x_i \partial x_k$, where F is any scalar function, is a tensor of the second order. Also, if we turn the axes through $\frac{1}{2}\pi$, the matrix of the transformation is $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$, and therefore the matrix

$$G_{ik} = \begin{pmatrix} \frac{\partial^2 F}{\partial x_2^2} & -\frac{\partial^2 F}{\partial x_1 \partial x_2} \\ -\frac{\partial^2 F}{\partial x_1 \partial x_2} & \frac{\partial^2 F}{\partial x_1^2} \end{pmatrix} \dots \dots \dots (1)$$

of the components of $\partial^2 F / \partial x_i \partial x_k$, referred to the new axes, is another tensor. This fact is directly related to the expression of the stress components in terms of the χ function in problems of plane stress.

If ξ_1 and ξ_2 are in the directions of principal axes of curvature, and the curvatures in these directions are κ_1 and κ_2 , the couples per unit length tending to produce these curvatures are

$$M_1 = D(\kappa_1 + \sigma \kappa_2); \quad M_2 = D(\kappa_2 + \sigma \kappa_1), \quad \dots \dots \dots (2)$$

and the energy per unit area is

$$\begin{aligned} W &= \frac{1}{2} D (\kappa_1^2 + 2\sigma \kappa_1 \kappa_2 + \kappa_2^2) \\ &= \frac{1}{2} D \{ (\kappa_1 + \kappa_2)^2 - 2(1 - \sigma) \kappa_1 \kappa_2 \}. \quad \dots \dots \dots (3) \end{aligned}$$

But this is the form taken, for principal axes, by the scalar

$$\frac{1}{2} D \left[\left(\frac{\partial^2 w}{\partial x_1^2} + \frac{\partial^2 w}{\partial x_2^2} \right)^2 - 2(1 - \sigma) \left\{ \frac{\partial^2 w}{\partial x_1^2} \frac{\partial^2 w}{\partial x_2^2} - \left(\frac{\partial^2 w}{\partial x_1 \partial x_2} \right)^2 \right\} \right], \quad \dots (4)$$

* Communicated by the Author.

the last term being the determinant of the elements of the tensor $\partial^2 w / \partial x_i \partial x_k$, and therefore, with any axes, will transform to this scalar. Now, for any axes, we can write

$$\kappa_{ik} = \frac{\partial^2 w}{\partial x_i \partial x_k} \quad \dots \quad (5)$$

$$\left. \begin{aligned} M_{11} &= D(\kappa_{11} + \sigma \kappa_{22}) \\ M_{22} &= D(\kappa_{22} + \sigma \kappa_{11}) \\ M_{12} &= M_{21} = D(1 - \sigma) \kappa_{12} \end{aligned} \right\} \quad \dots \quad (6)$$

and M_{ik} is a tensor, being the sum of $D\kappa_{ik}$ and $D\sigma$ times the tensor formed from w as in (1). For small variations $\delta w = w'$,

$$\delta W = M_{ik} \frac{\partial^2 w'}{\partial x_i \partial x_k} \quad \dots \quad (7)$$

and the variation in the total elastic energy of the plate is

$$\begin{aligned} \delta V &= \iint \delta W dS, \\ &= \int l_i M_{ik} \frac{\partial w'}{\partial x_k} ds - \iint \frac{\partial}{\partial x_k} \left(\frac{\partial M_{ik}}{\partial x_i} w' \right) dS + \iint \frac{\partial^2 M_{ik}}{\partial x_i \partial x_k} w' dS, \\ &= \int l_i M_{ik} \frac{\partial w'}{\partial x_k} ds - \int l_k \frac{\partial M_{ik}}{\partial x_i} w' ds + D \iint \nabla_1^4 w \cdot w' dS, \quad \dots \quad (8) \end{aligned}$$

l_i being direction cosines of the outward normal to the boundary, by two applications of Green's theorem in two dimensions. The integrands are all scalars and therefore preserve their forms when the axes are rotated in any manner. With each point of the boundary we associate a pair of Cartesian axes, y_1 along the outward normal and y_2 along the tangent. Denote the components of M_{ik} with regard to these by M_{nn} , M_{ns} , M_{ss} to avoid confusion of suffixes. Then $l_n = 1$, $l_s = 0$, and the boundary integrals become

$$\begin{aligned} &\int M_{nn} \frac{\partial w'}{\partial n} ds + \int \left(M_{ns} \frac{\partial w'}{\partial s} - \frac{\partial M_{nn}}{\partial n} w' - \frac{\partial M_{ns}}{\partial s} w' \right) ds \\ &= \int M_{nn} \frac{\partial w'}{\partial n} ds - \int w' \left(\frac{\partial M_{nn}}{\partial n} + 2 \frac{\partial M_{ns}}{\partial s} \right) ds \quad \dots \quad (9) \end{aligned}$$

on integration by parts.

For equilibrium the increase in elastic energy for all forms of w' must be supplied by the external forces. If these consist of a normal force Z per unit area over the plate, a couple M per unit length about the tangent to the boundary, and a force N per unit length normal to the plate along the boundary, we have therefore the differential equation

$$D \nabla_1^4 w = Z, \quad \dots \quad (10)$$

and the boundary conditions

$$M_{nn} = M. \quad (11)$$

$$\frac{\partial M_{nn}}{\partial n} + 2 \frac{\partial M_{ns}}{\partial s} = -N. \quad (12)$$

The components M_{nn} and M_{ns} involve second derivatives of w , and for application must be transformed from the Cartesian co-ordinates y_1 and y_2 to the curvilinear ones n and s . If the normal at a neighbouring point is at an angle θ to the axis of y_1 and we measure s along the are, we have

$$\frac{\partial}{\partial y_1} = \cos \theta \frac{\partial}{\partial n} - \sin \theta \frac{\partial}{\partial s} = \frac{\partial}{\partial n} - \theta \frac{\partial}{\partial s} + O(\theta^2), \quad (13)$$

$$\frac{\partial}{\partial y_2} = \sin \theta \frac{\partial}{\partial n} + \cos \theta \frac{\partial}{\partial s} = \frac{\partial}{\partial s} + \theta \frac{\partial}{\partial n} + O(\theta^2), \quad (14)$$

$$\frac{\partial^2 w}{\partial y_1^2} = \frac{\partial^2 w}{\partial n^2} + O(\theta); \quad \frac{\partial^2 w}{\partial y_2^2} = \frac{\partial^2 w}{\partial s^2} + \frac{1}{\rho} \frac{\partial w}{\partial n}; \quad \frac{\partial^2 w}{\partial y_1 \partial y_2} = \frac{\partial^2 w}{\partial n \partial s}, \quad (15)$$

and finally

$$M = D \left\{ \frac{\partial^2 w}{\partial n^2} + \sigma \left(\frac{\partial^2 w}{\partial s^2} + \frac{1}{\rho} \frac{\partial w}{\partial n} \right) \right\}, \quad (16)$$

$$-N = D \frac{\partial}{\partial n} \left\{ \frac{\partial^2 w}{\partial n^2} + (2 - \sigma) \left(\frac{\partial^2 w}{\partial s^2} + \frac{1}{\rho} \frac{\partial w}{\partial n} \right) \right\}. \quad (17)$$

The method is equivalent to that of Rayleigh, but is much abbreviated by the use of tensor notation. In particular the approach to (9), which is immediate here because the integrands in (8) are all scalars, involves a rather elaborate transformation of second derivatives when tensors are not used.

If there are edge thrusts in the plane, as in problems of elastic stability, the extension is simple. Principal thrusts P_1 and P_2 do work in a displacement w equal to

$$\frac{1}{2} P_1 \left(\frac{\partial w}{\partial x_1} \right)^2 + \frac{1}{2} P_2 \left(\frac{\partial w}{\partial x_2} \right)^2 = \frac{1}{2} P_{ik} \frac{\partial w}{\partial x_i} \frac{\partial w}{\partial x_k}, \quad (18)$$

which, being scalar, is correct for any axes. The contribution to δV in a varied displacement is then

$$\iint P_{ik} \frac{\partial w}{\partial x_i} \frac{\partial w'}{\partial x_k} dS = \int l_k P_{ik} \frac{\partial w}{\partial x_i} w' ds - \iint w' \frac{\partial}{\partial x_k} \left(P_{ik} \frac{\partial w}{\partial x_i} \right) dS, \quad (19)$$

and we must add $-\frac{\partial}{\partial x_k} \left(P_{ik} \frac{\partial w}{\partial x_i} \right)$ to the right of (10), $-P_{ni} \frac{\partial w}{\partial x_i}$ to the

left of (17), 0 to (16). If the boundary thrusts in the plane consist of a normal force P and a tangential force S per unit length,

$$\begin{aligned} P_{ni} \frac{\partial w}{\partial x_i} &= P_{nn} \frac{\partial w}{\partial n} + P_{ns} \frac{\partial w}{\partial s} \\ &= P \frac{\partial w}{\partial n} + S \frac{\partial w}{\partial s} \dots \dots \dots (20) \end{aligned}$$

Couples about the normal in the plane can be supplied not only by forces normal to the medial plane but by forces tangential to the edge and parallel to the medial plane. The resultant force along any tangent gives S . But moments H per unit length of the boundary about the normal in the medial plane will do work in a small displacement equal to

$$\int H \frac{\partial w'}{\partial s} ds = - \int \frac{\partial H}{\partial s} w' ds, \dots \dots \dots (21)$$

which must be added to the work done by the forces N , namely $\int N w' ds$. Hence, when such couples exist, they can be taken into account by replacing N in (12) and (17) by $N - \partial H / \partial s$.

This operation results in eliminating both the arbitrary constants and the roots and we deduce the following theorem.

If $x^{(0)}, x^{(1)}, \dots, x^{(m-1)}$ are m successive iterates to the value of any unknown and if

$$p_m E^m + p_{m-1} E^{m-1} + \dots + p_2 E^2 + p_1 E = 0 \quad (2.11)$$

is the expansion of the determinantal equation

$$\begin{vmatrix} E a_{11} & a_{12} & \dots & a_{1m} \\ E a_{21} & E a_{22} & \dots & a_{2m} \\ \dots & \dots & \dots & \dots \\ E a_{m1} & E a_{m2} & \dots & a_{mm} \end{vmatrix} = 0 \quad (2.12)$$

in powers of E , then the true value of the unknown is

$$\frac{p_m x^{(m-1)} + p_{m-1} x^{(m-2)} + \dots + p_2 x^{(1)} + p_1 x^{(0)}}{p_m + p_{m-1} + \dots + p_2 + p_1} \quad (2.13)$$

It may be noted that if

$$p_m + p_{m-1} + \dots + p_1 = 0, \quad (2.14)$$

the original equations (2.1) are either inconsistent or not independent.

3. An Example.

A simple example will illustrate the theorem. To solve

$$x + 4y - 10z = 1, \quad (3.1)$$

$$2x + 3y + 8z = 20, \quad (3.2)$$

$$3x + 5y + 2z = 21, \quad (3.3)$$

for which $x=3, y=2, z=1$.

If we take first approximations $y^{(0)}=0, z^{(0)}=0$.

$$(1) \text{ gives } x^{(1)} = 1 - 4y^{(0)} + 10z^{(0)}, \text{ hence } x^{(1)} = 1. \quad (3.4)$$

$$(2) \text{ gives } 3y^{(1)} = 20 - 2x^{(1)} - 8z^{(0)}, \text{ hence } y^{(1)} = 6. \quad (3.5)$$

$$(3) \text{ gives } 2z^{(1)} = 21 - 3x^{(1)} - 5y^{(1)}, \text{ hence } z^{(1)} = -6. \quad (3.6)$$

$$(1) \text{ gives } x^{(2)} = 1 - 4y^{(1)} + 10z^{(1)}, \text{ hence } x^{(2)} = -83. \quad (3.7)$$

$$(2) \text{ gives } 3y^{(2)} = 20 - 2x^{(2)} - 8z^{(1)}, \text{ hence } y^{(2)} = 78. \quad (3.8)$$

$$(3) \text{ gives } 2z^{(2)} = 21 - 3x^{(2)} - 5y^{(2)}, \text{ hence } z^{(2)} = -60. \quad (3.9)$$

$$(1) \text{ gives } x^{(3)} = 1 - 4y^{(2)} + 10z^{(2)}, \text{ hence } x^{(3)} = -911. \quad (3.10)$$

The equation for E is

$$\begin{vmatrix} E & 4 & -10 \\ 2E & 3E & 8 \\ 3E & 5E & 2E \end{vmatrix} = 0. \quad (3.11)$$

i. e.,
$$E^3 - 11E^2 + 16E = 0. \quad (3.12)$$

If we take first approximations $z^{(0)}=0, y^{(0)}=0,$

(3) gives $x^{(1)}=7$	(4) gives $z^{(1)}=0.75$	(5) gives $y^{(1)}=0.375$
(3) gives $x^{(2)}=5.875$	(4) gives $z^{(2)}=0.89$	(5) gives $y^{(2)}=1.01$
(3) gives $x^{(3)}=4.72$	(4) gives $z^{(3)}=0.94$	(5) gives $y^{(3)}=1.42$
(3) gives $x^{(4)}=4.01$	(4) gives $z^{(4)}=0.96$	(5) gives $y^{(4)}=1.65$
(3) gives $x^{(5)}=3.61$	(4) gives $z^{(5)}=0.98$	(5) gives $y^{(5)}=1.80$
(3) gives $x^{(6)}=3.35$	(4) gives $z^{(6)}=0.99$	(5) gives $y^{(6)}=1.89$
(3) gives $x^{(7)}=3.19$	(4) gives $z^{(7)}=0.99$	(5) gives $y^{(7)}=1.93$
(3) gives $x^{(8)}=3.12$	(4) gives $z^{(8)}=1.00$	(5) gives $y^{(8)}=1.97$
(3) gives $x^{(9)}=3.05$	(4) gives $z^{(9)}=1.00$	(5) gives $y^{(9)}=1.99$
(3) gives $x^{(10)}=3.02$	(4) gives $z^{(10)}=1.00$	(5) gives $y^{(10)}=2.00$

5. Method of Finding the Unknowns Approximately.

In the example just given the successive iterates diverged when we found them in one order and converged when we used another order of finding them. In general, when we have m unknowns, the n th iterate to any unknown x is-

$$x+A_1e_1^n+A_2e_2^n+\ldots+A_{m-1}e_{m-1}^n, \quad \text{.} \quad (5.1)$$

in which some of the e 's are less and some greater than unity. After a few iterations the values of e_p much less than unity will hardly affect the results. In that case the n th iterate will be

$$x+A_1e_1^n+A_2e_2^n+\ldots+A_pe_p^n, \quad \text{.} \quad (5.2)$$

where $p \leq m$, and if the order of finding the unknowns is chosen carefully, p will often be 0, 1, 2, or 3, even with equations involving a large number of unknowns. At the stage at which the above expression begins to represent the iterate approximately, we shall change our notation and call the iterate $y^{(0)}$, and the subsequent ones $y^{(1)}, y^{(2)}$, etc. Hence

$$\begin{aligned} y^{(0)} &= x+B_1+B_2+\ldots+B_p, \\ y^{(1)} &= x+B_1e_1+B_2e_2+\ldots+B_pe_p, \\ y^{(2)} &= x+B_1e_1^2+B_2e_2^2+\ldots+B_pe_p^2, \\ &\text{.} \\ y^{(p)} &= x+B_1e_1^p+B_2e_2^p+\ldots+B_pe_p^p. \end{aligned} \quad \text{.} \quad (5.3)$$

Let the equation whose roots are $e_1, e_2, \ldots e_p$ be

$$q_pE^p+q_{p-1}E^{p-1}+\ldots+q_1E+q_0=0. \quad \text{.} \quad (5.4)$$

Multiply the first equation by q_0 , the second by q_1 , etc., and add all resulting equations. We get

$$q_0(x-y^{(0)}+q_1(x-y^{(1)})+\ldots+q_p(x-y^{(p)}))=0.$$

Similarly,

$$\left. \begin{aligned} q_0(x-y^{(1)})+q_1(x-y^{(2)})+\dots+q_p(x-y^{(p+1)})=0, \\ \vdots \\ q_0(x-y^{(p)})+q_1(x-y^{(p+1)})+\dots+q_n(x-y^{(2p)})=0. \end{aligned} \right\} \quad . \quad (5.5)$$

Hence

$$\begin{vmatrix} x-y^{(0)} & x-y^{(1)} & \dots & x-y^{(p)} \\ x-y^{(1)} & x-y^{(2)} & \dots & x-y^{(p+1)} \\ \vdots & \vdots & \ddots & \vdots \\ x-y^{(p)} & x-y^{(p+1)} & \dots & x-y^{(2p)} \end{vmatrix} = 0. \quad (5.6)$$

Denote $y^{(1)} - y^{(0)}$ by $\Delta y^{(0)}$, $y^{(2)} - y^{(1)}$ by $\Delta y^{(1)}$, etc., $\Delta y^{(1)} - \Delta y^{(0)}$ by $\Delta^2 y^{(0)}$, $\Delta y^{(2)} - \Delta y^{(1)}$ by $\Delta^2 y^{(1)}$, etc.

In the above determinant subtract the p th column from the $(p+1)$ th, then subtract the $(p-1)$ th column from the p th, then subtract the $(p-2)$ th column from the $(p-1)$ th, etc.. Then perform an exactly similar operation on the rows. We get on solving for x and rearranging the terms

$$x = \frac{\begin{vmatrix} y^{(0)} & y^{(1)} & \dots & y^{(p)} \\ y^{(1)} & y^{(2)} & \dots & y^{(p+1)} \\ \vdots & \vdots & \ddots & \vdots \\ y^{(p)} & y^{(p+1)} & \dots & y^{(2p)} \end{vmatrix}}{\begin{vmatrix} \Delta^2 y^{(0)} & \Delta^2 y^{(1)} & \dots & \Delta^2 y^{(p-1)} \\ \Delta^2 y^{(1)} & \Delta^2 y^{(2)} & \dots & \Delta^2 y^{(p)} \\ \vdots & \vdots & \ddots & \vdots \\ \Delta^2 y^{(p-1)} & \Delta^2 y^{(p)} & \dots & \Delta^2 y^{(2p-2)} \end{vmatrix}} \cdot \dots \cdot \dots \quad (5.7)$$

Another way of writing this result is

$$x = \frac{\begin{vmatrix} y^{(0)} \Delta y^{(0)} & \Delta^2 y^{(0)} & \dots & \Delta^p y^{(0)} \\ \Delta y^{(0)} & \Delta^2 y^{(0)} & \dots & \Delta^{p+1} y^{(0)} \\ \dots & \dots & \dots & \dots \\ \Delta^p y^{(0)} \Delta^{(p+1)} y^{(0)} & \Delta^{(p+2)} y^{(0)} & \dots & \Delta^{(2p)} y^{(0)} \end{vmatrix}}{\begin{vmatrix} \Delta^2 y^{(0)} & \Delta^3 y^{(0)} & \dots & \Delta^{(p+1)} y^{(0)} \\ \Delta^3 y^{(0)} & \Delta^4 y^{(0)} & \dots & \Delta^{(p+2)} y^{(0)} \\ \dots & \dots & \dots & \dots \\ \Delta^{(p+1)} y^{(0)} \Delta^{(p+2)} y^{(0)} & \dots & \dots & \Delta^{(2p)} y^{(0)} \end{vmatrix}} \quad (5.8)$$

The above theorems are, of course, accurate if we take all of the roots of the "E" equation into account, *i. e.*, if $p=m$, but usually it only pays to use them when we may neglect a number of the roots. To illustrate (5.8) we will apply it to the problem previously solved. In paragraph 4 iterates were found for the unknowns, the roots of the corresponding determinantal equation being 0.58 and 0.11. After a few iterations we may expect that the latter root will hardly affect the results. Hence $p=1$ and

$$x \sim \frac{\begin{vmatrix} x^{(3)} & x^{(4)} \\ x^{(4)} & x^{(4)} \end{vmatrix}}{\Delta^2 x^{(3)}} = \frac{0.96}{0.31} \sim 3.10. \quad (5.9)$$

Similarly for the other unknowns.

In order to make use of the results just given it is best to arrange the equations so that r is small and then the evaluation of determinants of high order can be avoided. This is done by arranging the equations so that, as far as possible, the largest coefficient in any equation lies on the leading diagonal. If the coefficients in any equation are nearly equal, large diagonal terms can often be arranged by suitable additions and subtractions of the original equations.

6. A Criterion showing when the Method of Approximation may be used.

It is of importance to know when the approximate formula may be used. Fortunately a simple criterion can be found. Eliminating $q_0, q_1, q_2, \dots$ and x from the equations (5.5), we get after a simple transformation

$$\begin{vmatrix} \Delta y^{(0)} & \Delta y^{(1)} & \Delta y^{(2)} & \dots & \Delta y^{(p)} \\ \Delta y^{(1)} & \Delta y^{(2)} & \Delta y^{(3)} & \dots & \Delta y^{(p+1)} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \Delta y^{(p)} & \Delta y^{(p+1)} & \Delta y^{(p+2)} & \dots & \Delta y^{(2p)} \end{vmatrix} = 0, \quad (5.10)$$

which is the required criterion. In a numerical example, the determinant will generally not vanish, but will be small compared to its constituent terms.

Some Particular Cases.

Some particular cases of the theorems of the last two paragraphs are worth noting.

If $p=1$, equation (5.7) becomes

$$x = y^{(0)} - \frac{\{\Delta y^{(0)}\}^2}{\Delta^2 y^{(0)}}, \quad (5.11)$$

while the criterion for its use becomes

$$\frac{\Delta y^{(0)}}{\Delta y^{(1)}} = \frac{\Delta y^{(1)}}{\Delta y^{(2)}} \quad (5.12)$$

If the corresponding root of the determinantal equation is known, we have the relation

$$x = y^{(0)} - \frac{\Delta y^{(0)}}{e - 1} \quad (5.13)$$

If $p=2$, equation (5.7) becomes

$$x = y^{(0)} - \frac{\{\Delta y^{(0)}\}^2}{\Delta^2 y^{(0)}} - \frac{\{\Delta y^{(0)} \Delta y^{(2)} - [\Delta y^{(1)}]^2\}^2}{\Delta^2 y^{(0)} \{\Delta^2 y^{(0)} \Delta^2 y^{(2)} - [\Delta^2 y^{(1)}]^2\}} \quad (5.14)$$

6. Another Method of Approximation.

In the previous paragraph the finding of the unknown made use of the iterates to only one unknown. A method of making use of the iterates to several of the unknowns, which is usually more useful, will now be given.

In paragraph 5 it was shown that

$$x_r = \frac{q_0 y_r^{(0)} + q_1 y_r^{(1)} + \dots + q_p y_r^{(p)}}{q_0 + q_1 + \dots + q_p} \quad (6.1)$$

Subtracting the second equation in (5.6) from the first we get

$$q_0 \Delta y_r^{(0)} + q_1 \Delta y_r^{(1)} + \dots + q_p \Delta y_r^{(p)} = 0 \quad (6.2)$$

Now, if $x_r^{(n)}$ depends, approximately, on the roots $e_1, e_2, \dots, e_p$, only, then the iterates to other unknowns y_e , etc., will usually depend on these same roots only. Hence

$$q_0 \Delta y_e^{(0)} + q_1 \Delta y_e^{(1)} + \dots + q_p \Delta y_e^{(n)} = 0 \quad (6.3)$$

Now if we have p equations of this type, we can solve for q_0, q_1 , etc., and substituting in (6.1) we can find x_r and similarly the other unknowns.

It is the first differences of the iterates which are first obtained in Morris's ⁽²⁾ tabular method, and the equations (6.3) can therefore be easily written down.

A criterion telling us if this theorem can be used is clearly found by eliminating $q_0, q_1, \dots, q_p$ from equations of the form (6.4). In practice, however, it is found more convenient to solve p of the equations for q_0, q_1, q_2 , etc., and to find if the other equations are satisfied approximately.

7. Application of Previous Theory to an Example.

As an application of the foregoing theory we shall solve a set of simultaneous equations, which occurred in an aeronautical problem, and is taken from a paper by Winny ⁽⁵⁾. The equations are given in tabular form in Table I, A, B, C, etc. being the unknowns. The elements of

TABLE I.

Coefficients of								Constants.	
A.	B.	C.	D.	E.	F.	G.	H.		
74.7	126.6	24.95	450	33.5	57.6	107	186	-570	(7.1)
93.1	83.0	31.1	-122.5	42.6	41.5	15.35	59.2	-367	(7.2)
62.4	-14.7	-35.1	37.2	22.75	-5.77	-12.85	12.9	-169	(7.3)
-18.8	-15.1	-15.3	-14.7	5.71	-5.92	5.99	-5.54	-18.6	(7.4)
-0.269	-0.902	-2.33	-4.82	0.605	2.84	7.85	16.5	0	(7.5)
-0.499	-0.962	-0.499	2.410	0.905	2.58	1.48	-7.33	0	(7.6)
-0.437	0.237	0.835	-1.145	0.870	-0.765	-3.16	4.58	0	(7.7)
-0.210	0.445	-0.695	0.870	0.304	-1.14	2.269	-3.17	0	(7.8)

TABLE II.

Coefficients of									Constants.
A.	B.	C.	D.	E.	F.	G.	H.		
1	-.2356	-.5625	.5962	.3646	-.0925	-.2059	.2067	-2.7083	
.4903	1	.0044	-.8224	.2754	.4712	1.0298	.6208	-2.9134	
-.0741	-.3778	1	-1.0037	-.0694	-.1480	.3805	-.3644	1.3975	
.1660	.2813	.0554	1	.0744	.1280	.2378	.4133	-1.2667	
-.5368	-.0146	.1614	.3357	1	-.0305	-.8718	-.4544	0	
-.0163	-.3811	.1833	.1379	.0611	1	-.6292	-.2037	0	
-.0687	.0689	-.3024	-.0245	.1107	-.1490	1	.0339	0	
-.0163	.0547	-.1412	-.2921	.0367	.1721	.4758	1	0	
-.0559	-.7134	-.6015	-1.0730	.8536	.3513	-.5832	.2522		
Check sum (S)									

the table under each letter are the coefficients of that unknown in the various equations. The resulting terms in each row are to be added, and together with the constant equated to zero, forming the eight equations (7.1), (7.2), . . . (7.8).

TABLE III.

A.	B.	C.	D.	E.	F.	G.	H.	Check sums (C).
2.7083	2.9134	-1.3975	1.2667	0	0	0	0	
	-1.3279	.2007	- .4496	1.4538	.0441	-.1861	.0441	.1513
- .3736	1.5855	.5990	- .4460	.0231	.6043	- .1092	.0867	1.1315
- .3363	.6026	- .5978	.0331	-.0965	-.1096	- .1808	- .0844	- .3597
- .2409	.3323	.4056	.4041	- .1357	-.0557	.0099	.1181	.4336
- .5242	- .3959	.0998	- .1070	1.4377	- .0878	- .1592	- .0528	- 1.2271
.0568	- .2895	.0909	- .0786	-.0187	.6143	.0915	- .1057	- .2159
- .0333	.0048	.0615	.0385	- .1410	- .1018	- .1617	.0769	- .0944
- .0171	- .0515	.0302	- .0343	.0377	.0169	- .0028	.0829	- .0209
- .7214	.3537	- .0534	.1198	- .3872	- .6118	- .0496	- .0118	- .0403
- .0102	- .0435	- .0164	.0122	- .0006	- .0166	.0030	- .0024	- .0310
.3477	- .0027	.6182	- .0342	- .0998	- .1133	.1869	.0873	.3719
.0498	- .0688	- .0839	- .0836	.0281	.0115	- .0020	- .0244	- .0897
.1984	.1498	- .0378	.0405	- .5441	.0332	.0602	- .0200	.4643
- .0168	.0857	- .0269	.0233	- .0055	- .1819	- .0271	.0313	.0640
.0347	- .0050	- .0642	- .0401	.1470	.1061	.1686	.0802	.0983
- .0041	- .0123	.0072	- .0082	.0090	.0040	- .0007	.0198	- .0051
- .5995	- .2939	.0444	- .0995	.3218	.0098	.0412	.0098	1.0336
- .0347	- .1472	- .0556	.0414	- .0021	- .0561	.0101	- .0081	- .1051
- .1220	.0010	- .2168	.0120	.0350	.0397	- .0656	- .0306	- .1305
.0182	- .0252	.0307	- .0306	.0103	.0042	- .0007	- .0089	- .0328
.1880	- .1420	.0358	- .0384	.5155	- .0315	- .0571	- .0189	- .4401
.0070	- .0359	.0113	- .0098	.0023	.0762	.0114	- .0131	- .0268
- .0126	.0018	.0234	.0146	- .0535	- .0386	- .0614	.0292	- .0357
.0084	.0252	- .0148	.0168	- .0184	- .0083	.0014	- .0406	- .0103
- .3237	.1587	- .0240	.0537	- .1738	- .0053	- .0232	- .0053	- .0181
- .0039	- .0164	- .0062	.0046	- .0002	- .0062	.0111	- .0009	- .0117
- .0029	.0000	- .0052	.0003	.0008	.0010	- .0016	- .0007	- .0031
- .0249	.0344	.0420	.0418	- .0140	- .0058	.0010	.0122	.0449
.0936	.0707	- .0178	.0191	- .2568	.0157	.0284	.0094	.2191
- .0044	.0224	- .0070	.0061	- .0014	- .0475	- .0071	.0032	.0168
.0002	.0000	- .0004	- .0002	.0009	.0006	.0010	- .0005	.0006
- .0046	- .0139	.0082	- .0093	.0102	.0046	- .0008	.0224	- .0056
.0531	- .0260	.0039	- .0088	.0285	.0009	.0036	.0009	.0031
.0206	- .0876	.0331	- .0246	.0013	.0334	- .0060	.0048	.0626
.0349	- .0003	.0620	- .0034	- .0100	- .0114	.0187	.0088	.0373
.0126			- .0211	.0071	.0029	- .0005	- .0062	
- .0133				.0366	- .0022	- .0041	- .0013	
.0027						.0288	- .0043	
.0031							- .0050	
.0011							- .0072	
.0619							- .0052	

Equation (1) has the coefficient of D greater than the other coefficients in the equation and will therefore be taken as the new fourth equation.

Equation (3) will be taken as the new first and (5) as the new eighth, (2) - 2(4) will be taken as the new second, - (3) + 3(4) as the new third,

(6)+2(7)+(8) as new fifth, (6)−2(8) as new sixth, (5)+5(8) as new seventh.

Dividing each of these equations by the coefficient of the unknown in the leading diagonal term, the equations resulting are those given in Table II.

The approximations are then found as explained by Morris ⁽²⁾, except that some additional calculations are made for checking purposes. The sum of the coefficients *S* of each unknown (*i. e.* the elements of any column), except for unity in the leading diagonal, is found and the number is placed in that column under the coefficients. The sum (*C.*) of the elements of any row, except for the actual iterate in that row, is also found. Then the product of *S* and the iterate should equal minus and check sum *C*. Thus the elements of each row in the calculation are checked as they are found. The results are given in Table III.

Hence the iterates to the unknowns are given by the following table :

TABLE IV.

	First.	Second.	Third.	Fourth.	Fifth.
A	2.7083	1.9869	2.5866	2.2627	2.3158
B	1.5855	1.5422	1.3948	1.3784	1.4661
C	−0.5978	0.0204	−0.01964	−0.2016	−0.1396
D	0.4041	0.3205	0.2899	0.3317	0.3106
E	1.4377	0.8936	1.4091	1.1523	1.1889
F	0.6143	0.4324	0.5086	0.4611	0.4899
G	−0.1617	0.0069	−0.0545	−0.0535	−0.0383
H	0.0829	0.1027	0.0621	0.0845	0.0793

It should be noted that we could have written down the iterates in Table IV. direct from Table II. as the iterates can be found directly with the aid of a calculating machine. This method is, however, more difficult to check.

We now take the differences of the iterates and see on how many roots of the "E" equation they depend. It is obvious that they depend on more than one, since

$$\frac{\Delta A^{(2)}}{\Delta A^{(3)}} \neq \frac{\Delta A^{(3)}}{\Delta A^{(4)}}.$$

If they depended on two, then

$$\begin{vmatrix} \Delta A^{(2)} & \Delta A^{(3)} & \Delta A^{(4)} \\ \Delta B^{(2)} & \Delta B^{(3)} & \Delta B^{(4)} \\ \Delta C^{(2)} & \Delta C^{(3)} & \Delta C^{(4)} \end{vmatrix}$$

would vanish. This is not so, but if the equations

$$\begin{aligned} 0.5995 q_0 &- 0.3237 q_1 &+ 0.0531 q_2 &+ 0.0619 q_3 = 0 \\ 0.6182 q_0 &- 0.2168 q_1 &- 0.0052 q_2 &+ 0.0620 q_3 = 0 \\ -0.5441 q_0 &+ 0.5155 q_1 &- 0.2568 q_2 &+ 0.0366 q_3 = 0 \end{aligned}$$

are solved, we get

$$q_0 : q_1 : q_2 : q_3 = 0.0361 : 0.2021 : 0.3831 : 0.3781.$$

The results satisfy approximately the equations

$$q_0 \Delta B^{(2)} + q_1 \Delta B^{(3)} + q_2 \Delta B^{(4)} + q_3 \Delta B^{(5)} = 0, \text{ etc.}$$

We can choose any three of eight equations to find the ratios $q_0 : q_1 : q_2 : q_3$. The three equations (7.10) were selected as they have the largest coefficients and will therefore yield the most accurate values to these auxiliary unknowns.

Making use of (6.1), we find

$$\begin{aligned} A &= 2.338 & B &= 1.421 & C &= -0.169 & D &= 0.315 \\ E &= 1.209 & F &= 0.481 & G &= -0.045 & H &= 0.78. \end{aligned}$$

When these values are substituted into the original equations (7.1), the remainders are

$$0.08, 0.20, 0.003, 0.04, 0.001, 0.001, 0.00001, 0.00002.$$

Hence the solutions found are approximately correct.

If we had evaluated the iterates to only three decimal places, found one iterate less to each of the unknowns, and solved for the ratios

$$q_0 : q_1 : q_2 : q_3,$$

we would have got

$$\begin{aligned} A &= 2.33 & B &= 1.41 & C &= -0.18 \\ D &= 0.31 & E &= 1.20 & F &= 0.47 \\ G &= -0.05 & H &= 0.08 \end{aligned}$$

By evaluating the iterates to five decimal places and finding the first six iterates to the unknowns, it was found that

$$\begin{aligned} A &= 2.3377 & B &= 1.4204 & C &= -0.1702 \\ D &= 0.3153 & E &= 1.2069 & F &= 0.4806 \\ G &= -0.0498 & H &= 0.0786 \end{aligned}$$

The writer is deeply indebted to Dr. W. G. Bickley for many valuable suggestions.

References.

- (1) Whittaker and Robinson, 'The Calculus of Observations,' Section 130.
- (2) J. Morris, 'Reports and Memoranda of the Aeronautical Research Committee No. 1711 ; and 'Journal of the Royal Aeronautical Society,' April 1935.
- (3) Southwell, 'Proceedings of the Royal Society,' A, cli. (1935), etc.
- (4) Milne-Thomson, 'Calculus of Finite Differences,' Cap. 8.
- (5) Winny, 'Reports and Memoranda of the Aeronautical Research Committee,' No. 1756.

XXXIII. *The Conduction of Heat in a Medium Generating Heat.*

By STEWART PATERSON, M.A., B.Sc.,
late Metcalfe Fellow, Glasgow University *.

[Received April 14, 1941.]

Introduction.

THE problem of thermal conduction is frequently complicated in practice by local generation or absorption of heat. This happens, for example, when heat energy is dissipated throughout the medium by the passage of electric current, or when it is generated or absorbed in a change of state such as fusion or chemical reaction. Certain particular cases have been discussed †, in which the rate of generation did not depend explicitly on the temperature, and the classical method of sources provides the basis for a more general treatment, under the same conditions: but when large temperature differences occur and when the rate of generation depends markedly upon temperature, as in most problems of chemical physics where heat conduction plays a part, such assumptions evidently cannot be made. It is still possible, however, to derive explicit solutions of a very general kind when the rate of generation is a linear function of temperature, provided the medium behaves as a solid with constant thermometric conductivity. Since these conditions may be taken as a first approximation to the truth in many cases of practical interest, the corresponding solutions will perhaps be useful.

PART I.—INFINITE SOLID.

The conduction of heat in an infinite homogeneous isotropic solid, in which heat is being generated or absorbed, may be represented as a rule by the equations:

$$\left. \begin{aligned} \frac{\partial \theta}{\partial t} &= \kappa \nabla^2 \theta + G(x, y, z, t; \theta); \quad t > 0, \\ \theta &\rightarrow \phi(x, y, z), \quad \text{when } t \rightarrow 0, \end{aligned} \right\} \quad \text{all } x, y, z,$$

where κ is the thermometric conductivity, θ the temperature at the point (x, y, z) at time t , and G some specified function of the independent

* Communicated by Prof. E. Taylor Jones.

† *E. g.*, Awbery, *Phil. Mag.* (7) iv. p. 629 (1927); Orstrand, *Wash. Acad. Sci. J.* xxii. p. 530 (1932).

variables x, y, z, t and also of θ . $\phi(x, y, z)$ represents the initial temperature, supposed given, and ∇^2 the Laplacian operator $\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$. We shall solve these equations in the case where

$$G(x, y, z, t; \theta) \equiv f(x, y, z, t) + g(t) \cdot \theta.$$

Consider first the equations :

$$\left. \begin{aligned} \frac{\partial \theta}{\partial t} &= \kappa \nabla^2 \theta + f(x, y, z, t), \quad t > 0, \\ \theta &\rightarrow \phi(x, y, z), \quad t \rightarrow 0, \end{aligned} \right\} \text{all } x, y, z. \quad (1)$$

The solution, which may be obtained by the method of sources, or otherwise, is :

$$\begin{aligned} \theta &= \frac{1}{(4\pi\kappa t)^{3/2}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \phi(\alpha, \beta, \gamma) e^{-\frac{\Sigma(x-\alpha)^2}{4\kappa t}} d\alpha d\beta d\gamma \\ &+ \int_0^t \frac{1}{[4\pi\kappa(t-\tau)]^{3/2}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(\alpha, \beta, \gamma, \tau) e^{-\frac{\Sigma(x-\alpha)^2}{4\kappa(t-\tau)}} d\tau d\alpha d\beta d\gamma. \end{aligned} \quad (2)$$

The first term, which will be denoted by θ_1 , represents the solution when there is no generation of heat.

Turning now to the proposed equations, viz.,

$$\left. \begin{aligned} \frac{\partial \theta}{\partial t} &= \kappa \nabla^2 \theta + f(x, y, z, t) + g(t) \cdot \theta, \quad t > 0, \\ \theta &\rightarrow \phi(x, y, z), \quad t \rightarrow 0, \end{aligned} \right\} \text{all } x, y, z, \quad (3)$$

$$\text{set } \theta = v(x, y, z, t) \cdot e^{\int_0^t g(\tau) d\tau}.$$

Then v must satisfy

$$\left. \begin{aligned} \frac{\partial v}{\partial t} &= \kappa \nabla^2 v + f(x, y, z, t) e^{-\int_0^t g(\tau) d\tau}, \quad t > 0, \\ v &\rightarrow \phi, \quad t \rightarrow 0, \end{aligned} \right\} \text{all } x, y, z, \quad (4)$$

and it can readily be seen that, if it does so, θ will satisfy (3).

But the solution of (4), by (2) above, is

$$v = \theta_1 + \int_0^t \frac{e^{-\int_0^\tau g(\tau) d\tau}}{[4\pi\kappa(t-\tau)]^{3/2}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(\alpha, \beta, \gamma, \tau) e^{-\frac{\Sigma(x-\alpha)^2}{4\kappa(t-\tau)}} d\tau d\alpha d\beta d\gamma. \quad (5)$$

Hence the solution of (3) is

$$\begin{aligned} \theta &= e^{\int_0^t g(\tau) d\tau} \left\{ \frac{1}{(4\pi\kappa t)^{3/2}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \phi(\alpha, \beta, \gamma) e^{-\frac{\Sigma(x-\alpha)^2}{4\kappa t}} d\alpha d\beta d\gamma \right. \\ &\quad \left. + \int_0^t \frac{e^{-\int_0^\tau g(\tau) d\tau}}{[4\pi\kappa(t-\tau)]^{3/2}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(\alpha, \beta, \gamma, \tau) e^{-\frac{\Sigma(x-\alpha)^2}{4\kappa(t-\tau)}} d\tau d\alpha d\beta d\gamma \right\}. \end{aligned} \quad (6)$$

The corresponding solutions for flow of heat in two and one dimensions can readily be deduced from (6). They contain one, or two, less space integrations and one, or two, less powers of $\frac{1}{(4\pi\kappa t)^{1/2}}$ and $\frac{1}{[4\pi\kappa(t-\tau)]^{1/2}}$.

The solution for the case of spherical symmetry can also be deduced from (6). It is formally simpler, however, to derive it, by a well-known method, from the solution for a semi-infinite solid whose bounding plane is maintained at zero temperature. This will be done in the sequel. The solution is

$$\theta = \frac{e^{\int_0^t g(\tau) d\tau}}{r} \left\{ \frac{1}{(4\pi\kappa t)^{1/2}} \int_0^\infty \rho \phi(\rho) \left[e^{-\frac{(r-\rho)^2}{4\kappa t}} - e^{-\frac{(r+\rho)^2}{4\kappa t}} \right] d\rho \right. \\ \left. + \int_0^t \frac{e^{-\int_0^\tau g(\tau) d\tau}}{[4\pi\kappa(t-\tau)]^{1/2}} \int_0^\infty \rho f(\rho, \tau) \left[e^{-\frac{(r-\rho)^2}{4\kappa(t-\tau)}} - e^{-\frac{(r+\rho)^2}{4\kappa(t-\tau)}} \right] d\tau d\rho \right\}, \quad (7)$$

$\phi(r)$ and $f(r, t)$ corresponding to the $\phi(x, y, z)$ and $f(x, y, z, t)$ of equations (3).

Various interesting results can be deduced for particular cases from (6) and (7). Thus, if f and g are constant in space and time, the solutions are

$$\theta = \frac{f}{g} (e^{gt} - 1) + \frac{e^{gt}}{(4\pi\kappa t)^{3/2}} \int_{-\infty}^\infty \int_{-\infty}^\infty \int_{-\infty}^\infty \phi(\alpha, \beta, \gamma) e^{-\frac{\Sigma(x-\alpha)^2}{4\kappa t}} d\alpha d\beta d\gamma, \quad (8)$$

$$\text{and} \quad \theta = \frac{f}{g} (e^{gt} - 1) + \frac{e^{gt}}{r(4\pi\kappa t)^{1/2}} \int_0^\infty \rho \phi(\rho) \left[e^{-\frac{(r-\rho)^2}{4\kappa t}} - e^{-\frac{(r+\rho)^2}{4\kappa t}} \right] d\rho. \quad (9)$$

PART II.—SEMI-INFINITE SOLID.

When the solid is in any way bounded, in addition to the equations (3), which will now apply only over a limited range of the space co-ordinates, we shall have a surface condition. When, as in most practical cases, this condition can be expressed by the vanishing of a linear function of the temperature and normal temperature gradient at the surface, the *general* problem can be reduced, by the method of Part I., to the corresponding *general* problem without generation of heat. For, if (R) denote the range of x, y, z comprised by the solid, and $\mathbf{n}$ the normal vector drawn inward from a point on the bounding surface, the equations will then be of the form :

$$\left. \begin{aligned} \frac{\partial \theta}{\partial t} &= \kappa \nabla^2 \theta + f(x, y, z, t) + g(t), \quad \theta, t > 0, \text{ in (R),} \\ \theta &\rightarrow \phi(x, y, z), \quad t \rightarrow 0, \text{ in (R),} \\ \psi_1(x, y, z, t) \frac{\partial \theta}{\partial n} - \psi_2(x, y, z, t) \theta + \psi_3(x, y, z, t) &\rightarrow 0, \quad n \rightarrow 0, t > 0 \end{aligned} \right\} \quad (10)$$

The term $g(t)\theta$ can be removed by the substitution $\theta = ve^{\int_0^t g(\tau) d\tau}$, without otherwise altering the form of the equations. We therefore consider :

$$\left. \begin{aligned} \frac{\partial v}{\partial t} &= \kappa \nabla^2 v + f'(x, y, z, t), \quad t > 0, \text{ in } (R), \\ v &\rightarrow \phi, \quad t \rightarrow 0, \text{ in } (R), \\ \psi_1 \frac{\partial v}{\partial n} - \psi_2 v + \psi_3' &\rightarrow 0, \quad n \rightarrow 0, \quad t > 0 \end{aligned} \right\}, \quad \dots \quad (11)$$

where

$$f' \equiv f \cdot e^{-\int_0^t g(\tau) d\tau}, \quad \psi_3' \equiv \psi_3 \cdot e^{-\int_0^t g(\tau) d\tau}.$$

Now, if

$$\left. \begin{aligned} \frac{\partial \theta_1}{\partial t} &= \kappa \nabla^2 \theta_1, \quad t > 0, \text{ in } (R), \\ \theta_1 &\rightarrow \phi, \quad t \rightarrow 0, \text{ in } (R), \\ \psi_1 \frac{\partial \theta_1}{\partial n} - \psi_2 \theta_1 + \psi_3' &\rightarrow 0, \quad n \rightarrow 0, \quad t > 0 \end{aligned} \right\}, \quad \dots \quad (12)$$

and $w(x, y, z, t, \lambda)$ satisfies

$$\left. \begin{aligned} \frac{\partial w}{\partial t} &= \kappa \nabla^2 w, \quad t > 0, \text{ in } (R), \\ w &\rightarrow f'(x, y, z, \lambda), \quad t \rightarrow 0, \text{ in } (R), \\ \psi_1 \frac{\partial w}{\partial n} - \psi_2 w &\rightarrow 0, \quad n \rightarrow 0, \quad t > 0 \end{aligned} \right\}, \quad \dots \quad (13)$$

the solution of (11) is $v = \theta_1 + \int_0^t w(x, y, z, t - \tau, \tau) d\tau$; for this satisfies

the first two equations of (11), as can be verified, and $\psi_1 \frac{\partial v}{\partial n} - \psi_2 v + \psi_3'$

$$\begin{aligned} &= \psi_1 \frac{\partial \theta_1}{\partial n} + \psi_1 \int_0^t \frac{\partial}{\partial n} w(x, y, z, t - \tau, \tau) d\tau - \psi_2 \theta_1 \\ &\quad - \psi_2 \int_0^t w(x, y, z, t - \tau, \tau) d\tau + \psi_3' \end{aligned}$$

$$= \left(\psi_1 \frac{\partial \theta_1}{\partial n} - \psi_2 \theta_1 + \psi_3' \right) + \int_0^t \left(\psi_1 \frac{\partial w}{\partial n} - \psi_2 w \right) d\tau \rightarrow 0 \text{ as } n \rightarrow 0, \quad t > 0.$$

Since (13) are of the form (12), we have therefore reduced the general solution of (10) to that of (12), i. e., of the corresponding *general* problem without generation of heat.

The three cases commonly discussed are those in which

- (a) the surface temperature is prescribed,
- (b) there is no heat loss through the surface, and
- (c) there is a heat loss by radiation into a medium at temperature $\psi(x, y, z, t)$.

The first case corresponds to $\psi_1(x, y, z, t) \equiv 0$, the second to

$$\psi_2(x, y, z, t) \equiv 0 \equiv \psi_3(x, y, z, t),$$

the third to
$$\frac{\psi_2(x, y, z, t)}{\psi_1(x, y, z, t)} \equiv \frac{\psi_3(z, y, x, t)}{\psi_1(x, y, z, t)\psi(x, y, z, t)} = h,$$

hk being the heat-transfer coefficient of the surface, where k is the thermal conductivity of the solid. The appropriate forms of (12) and (13) are obvious. We conclude with solutions for these three cases in the semi-infinite solids defined by $x > 0$ and $r > a$.

(a) *Prescribed surface temperature* : (1) $x > 0$.

The equations (10) become

$$\left. \begin{aligned} \frac{\partial \theta}{\partial t} &= \kappa \nabla^2 \theta + f(x, y, z, t) + g(t)\theta, \quad t > 0, x > 0, \\ \theta &\rightarrow \phi(x, y, z), \quad t \rightarrow 0, x > 0, \\ \theta &\rightarrow \psi(y, z, t), \quad x \rightarrow 0+, t > 0 \end{aligned} \right\} \text{all } y, z, \dots \quad (14)$$

and the solution is therefore $\theta = e^{\int_0^t g(\tau) d\tau} \left[\theta_1 + \int_0^t w(x, y, z, t - \tau, \tau) d\tau \right]$, where

$$\left. \begin{aligned} \frac{\partial \theta_1}{\partial t} &= \kappa \nabla^2 \theta_1, \quad t > 0, x > 0, \\ \theta_1 &\rightarrow \phi, \quad t \rightarrow 0, x > 0, \\ \theta_1 &\rightarrow \psi e^{-\int_0^t g(\tau) d\tau}, \quad x \rightarrow 0+, t > 0 \end{aligned} \right\} \text{all } y, z, \dots \quad (15)$$

and
$$\left. \begin{aligned} \frac{\partial}{\partial t} w(x, y, z, t, \lambda) &= \kappa \nabla^2 w, \quad t > 0, x > 0, \\ w &\rightarrow f(x, y, z, \lambda) e^{-\int_0^\lambda g(\tau) d\tau}, \quad t \rightarrow 0, x > 0, \\ w &\rightarrow 0, \quad x \rightarrow 0+, t > 0 \end{aligned} \right\} \text{all } y, z, \dots \quad (16)$$

From the known solutions of these equations we deduce

$$\begin{aligned} \theta &= e^{\int_0^t g(\tau) d\tau} \left\{ \frac{1}{(4\pi\kappa)^{3/2}} \int_0^t \frac{e^{-\int_0^\tau g(\tau) d\tau}}{(t-\tau)^{3/2}} \int_0^\infty \int_{-\infty}^\infty \int_{-\infty}^\infty f(\alpha, \beta, \gamma, \tau) e^{-\frac{\Sigma(x-\alpha)^2}{4\kappa(t-\tau)}} \left[1 - e^{-\frac{\alpha x}{\kappa(t-\tau)}} \right] \right. \\ &\quad \times d\tau d\alpha d\beta d\gamma \\ &\quad + \frac{1}{t^{3/2}} \int_0^\infty \int_{-\infty}^\infty \int_{-\infty}^\infty \phi(\alpha, \beta, \gamma) e^{-\frac{\Sigma(x-\alpha)^2}{4\kappa t}} \left[1 - e^{-\frac{\alpha x}{\kappa t}} \right] d\alpha d\beta d\gamma \\ &\quad \left. + x \int_0^t \frac{e^{-\int_0^\tau g(\tau) d\tau}}{(t-\tau)^{5/2}} \int_{-\infty}^\infty \int_{-\infty}^\infty \psi(\beta, \gamma, \tau) e^{-\frac{x^2 + (y-\beta)^2 + (z-\gamma)^2}{4\kappa(t-\tau)}} d\tau d\beta d\gamma \right\}. \quad (17) \end{aligned}$$

The corresponding solutions for flow in two and one dimensions can, as before, be readily deduced from the general solution.

$$(2) \ r > a.$$

The equations are :

$$\left. \begin{aligned} \frac{\partial \theta}{\partial t} &= \kappa \frac{1}{r} \frac{\partial^2}{\partial r^2} (r\theta) + f(r, t) + g(t)\theta, \quad t > 0, \ r > a, \\ \theta &\rightarrow \phi(r), \quad t \rightarrow 0, \ r > a, \\ \theta &\rightarrow \psi(t), \quad r \rightarrow a+, \quad t > 0. \end{aligned} \right\} \dots (18)$$

Substituting $r\theta = u$, $r = a + x$, we have

$$\begin{aligned} \frac{\partial u}{\partial t} &= \kappa \frac{\partial^2 u}{\partial x^2} + (a+x)f(a+x, t) + g(t)u, \quad t > 0, \ x > 0, \\ u &\rightarrow (a+x)\phi(a+x), \quad t \rightarrow 0, \ x > 0, \\ u &\rightarrow a\psi(t), \quad x \rightarrow 0+, \quad t > 0, \end{aligned}$$

the solution of which can be obtained from (17). Reverting to the original variables, we get

$$\begin{aligned} \theta &= \frac{e^{\int_0^t g(\tau) d\tau}}{r(4\pi\kappa)^{1/2}} \left\{ \int_0^t e^{-\int_0^\tau g(\tau) d\tau} \int_a^\infty \rho f(\rho, \tau) \left[e^{-\frac{(r-\rho)^2}{4\kappa(t-\tau)}} - e^{-\frac{(r-2a+\rho)^2}{4\kappa(t-\tau)}} \right] d\tau d\rho \right. \\ &\quad \left. + \frac{1}{t^{1/2}} \int_a^\infty \rho \phi(\rho) \left[e^{-\frac{(r-\rho)^2}{4\kappa t}} - e^{-\frac{(r-2a+\rho)^2}{4\kappa t}} \right] d\rho + a(r-a) \int_0^t e^{\int_0^\tau g(\tau) d\tau - \frac{(r-a)^2}{4\kappa(t-\tau)}} \psi(\tau) d\tau \right\} \end{aligned} \quad (19)$$

The solution for an *infinite* solid with spherical symmetry of flow can also be deduced from (17). In this case, the equations are

$$\left. \begin{aligned} \frac{\partial \theta}{\partial t} &= \kappa \frac{1}{r} \frac{\partial^2}{\partial r^2} (r\theta) + f(r, t) + g(t)\theta, \quad t > 0, \ r > 0, \\ \theta &\rightarrow \phi(r), \quad t \rightarrow 0, \ r > 0 \end{aligned} \right\}, \dots (20)$$

which reduce, with $r\theta = u$, to

$$\begin{aligned} \frac{\partial u}{\partial t} &= \kappa \frac{\partial^2 u}{\partial r^2} + rf(r, t) + g(t)u, \quad t > 0, \ r > 0, \\ u &\rightarrow r\phi(r), \quad t \rightarrow 0, \ r > 0, \\ u &\rightarrow 0, \quad r \rightarrow 0+, \quad t > 0. \end{aligned}$$

The solution of this can be written down from (17), and leads to the expression for θ given in (7) above.

(b) *No surface loss*: (1) $x > 0$.

Equations (10) are now:

$$\left. \begin{aligned} \frac{\partial \theta}{\partial t} &= \kappa \nabla^2 \theta + f(x, y, z, t) + g(t)\theta, \quad t > 0, x > 0, \\ \theta &\rightarrow \phi(x, y, z), \quad t \rightarrow 0, x > 0, \\ \frac{\partial \theta}{\partial x} &\rightarrow 0, \quad x \rightarrow 0+, t > 0. \end{aligned} \right\} \text{all } y, z. \quad (21)$$

(12) and (13) take similar form, but without the heat generation terms. Their solutions can be written down by the familiar method of extending the solid with symmetrical temperature field beyond $x=0$; and the value of θ is found to be

$$\begin{aligned} \theta = & \frac{e^{\int_0^t g(\tau) d\tau}}{(4\pi\kappa)^{3/2}} \left\{ \int_0^t \frac{e^{-\int_0^\tau g(\tau) d\tau}}{(t-\tau)^{3/2}} \int_0^\infty \int_{-\infty}^\infty \int_{-\infty}^\infty f(\alpha, \beta, \gamma, \tau) e^{-\frac{\Sigma(x-\alpha)^2}{4\kappa(t-\tau)}} \left[1 + e^{-\frac{\alpha x}{\kappa(t-\tau)}} \right] \right. \\ & \times d\tau d\alpha d\beta d\gamma \\ & \left. + \frac{1}{t^{3/2}} \int_0^\infty \int_{-\infty}^\infty \int_{-\infty}^\infty \phi(\alpha, \beta, \gamma) e^{-\frac{\Sigma(x-\alpha)^2}{4\kappa t}} \left[1 + e^{-\frac{\alpha x}{\kappa t}} \right] d\alpha d\beta d\gamma \right\}. \quad (22) \end{aligned}$$

Solutions for two- and one-dimensional flow follow as before.

(2) $r > a$.

(10) becomes

$$\left. \begin{aligned} \frac{\partial \theta}{\partial t} &= \kappa \frac{1}{r} \frac{\partial^2}{\partial r^2} (r\theta) + f(r, t) + g(t)\theta, \quad t > 0, r > a, \\ \theta &\rightarrow \phi(r), \quad t \rightarrow 0, r > a, \\ \frac{\partial \theta}{\partial r} &\rightarrow 0, \quad r \rightarrow a+, t > 0 \end{aligned} \right\}, \quad (23)$$

which transforms, with $r\theta = u$, $r = a + x$, into

$$\left. \begin{aligned} \frac{\partial u}{\partial t} &= \kappa \frac{\partial^2 u}{\partial x^2} + (a+x)f(a+x, t) + g(t)u, \quad t > 0, x > 0, \\ u &\rightarrow (a+x)\phi(a+x), \quad t \rightarrow 0, x > 0, \\ \frac{\partial u}{\partial x} - \frac{u}{a} &\rightarrow 0, \quad x \rightarrow 0+, t > 0. \end{aligned} \right\}. \quad (24)$$

The solution of equations (24), which represent linear flow with radiation into a medium at zero temperature, can be obtained from case (c), equation (27) below. In terms of the original variables θ, r it is:

$$\theta = \frac{e^{\int_0^t g(\tau) d\tau}}{r(4\pi\kappa)^{1/2}} \left\{ \frac{1}{t^{1/2}} \int_a^\infty \rho \phi(\rho) \left[\left(e^{-\frac{(r-\rho)^2}{4\kappa t}} + e^{-\frac{(r-2a+\rho)^2}{4\kappa t}} \right) - \frac{2}{a} \int_0^\infty e^{-\frac{\xi}{a}} e^{-\frac{(r+\rho+\xi-2a)^2}{4\kappa t}} d\xi \right] d\rho \right.$$

$$+ \int_0^t \frac{e^{-\int_0^\tau g(\tau) d\tau}}{(t-\tau)^{3/2}} \int_a^\infty \rho f(\rho, \tau) \left[\left(e^{-\frac{(r-\rho)^2}{4K(t-\tau)}} + e^{-\frac{(r-2a+\rho)^2}{4K(t-\tau)}} \right) - \frac{2}{a} \int_0^\infty e^{-\frac{\xi}{a}} \frac{(r+\rho+\xi-2a)^2}{4K(t-\tau)} d\xi \right] \times d\tau d\rho \Big\} \dots (25)$$

(c) Radiation into medium at prescribed temperature : (1) $x > 0$.

Equations (10) become in this case

$$\left. \begin{aligned} \frac{\partial \theta}{\partial t} &= \kappa \nabla^2 \theta + f(x, y, z, t) + g(t)\theta, \quad t > 0, x > 0, \\ \theta &\rightarrow \phi(x, y, z), \quad t \rightarrow 0, x > 0, \\ \frac{\partial \theta}{\partial x} - h[\theta - \psi(y, z, t)] &\rightarrow 0, \quad x \rightarrow 0+, t > 0 \end{aligned} \right\} \text{all } y, z, \dots (26)$$

and equations (11), (12), (13) can be written down at once. The solution of (12), and hence also of (13), can be obtained by means of the appropriate Green's function *, and the solution of (26) finally derived in the form :

$$\begin{aligned} \theta &= \frac{e^{-\int_0^t g(\tau) d\tau}}{(4\pi\kappa)^{3/2}} \left\{ \int_0^t \frac{e^{-\int_0^\tau g(\tau) d\tau}}{(t-\tau)^{3/2}} \int_0^\infty \int_{-\infty}^\infty \int_{-\infty}^\infty f(\alpha, \beta, \gamma, \tau) \left[e^{-\frac{\Sigma(x-\alpha)^2}{4K(t-\tau)}} \left(1 + e^{-\frac{ax}{K(t-\tau)}} \right) \right. \right. \\ &\quad \left. \left. - 2h \int_0^\infty e^{-h\xi - \frac{(x+\alpha+\xi)^2 + (y-\beta)^2 + (z-\gamma)^2}{4K(t-\tau)}} d\xi \right] d\tau d\alpha d\beta d\gamma \right. \\ &+ \frac{1}{t^{3/2}} \int_0^\infty \int_{-\infty}^\infty \int_{-\infty}^\infty \phi(\alpha, \beta, \gamma) \left[e^{-\frac{\Sigma(x-\alpha)^2}{4Kt}} \left(1 + e^{-\frac{ax}{Kt}} \right) \right. \\ &\quad \left. \left. - 2h \int_0^\infty e^{-h\xi - \frac{(x+\alpha+\xi)^2 + (y-\beta)^2 + (z-\gamma)^2}{4Kt}} d\xi \right] d\alpha d\beta d\gamma \right. \\ &+ h \int_0^t \frac{e^{-\int_0^\tau g(\tau) d\tau}}{(t-\tau)^{5/2}} \int_{-\infty}^\infty \int_{-\infty}^\infty \int_0^\infty \psi(\beta, \gamma, \tau)(x+\xi) e^{-h\xi - \frac{(x+\xi)^2 + (y-\beta)^2 + (z-\gamma)^2}{4K(t-\tau)}} \\ &\quad \left. \times d\tau d\beta d\gamma d\xi \right\} \dots (27) \end{aligned}$$

Two- and one-dimensional solutions follow easily.

(2) $r > a$.

Equations (10) are now :

$$\left. \begin{aligned} \frac{\partial \theta}{\partial t} &= \kappa \frac{1}{r} \frac{\partial^2}{\partial r^2} (r\theta) + f(r, t) + g(t)\theta, \quad t > 0, r > a, \\ \theta &\rightarrow \phi(r), \quad t \rightarrow 0, r > a, \\ \frac{\partial \theta}{\partial r} - h[\theta - \psi(t)] &\rightarrow 0, \quad r \rightarrow a+, t > 0 \end{aligned} \right\} \dots (28)$$

* See Carslaw, 'Conduction of Heat in Solids,' 2nd ed., p. 184, for this function.

which become, with $r\theta=u$, $r=a+x$,

$$\begin{aligned}\frac{\partial u}{\partial t} &= \kappa \frac{\partial^2 u}{\partial x^2} + (a+x)f(a+x, t) + g(t)u, \quad t > 0, \quad x > 0, \\ u &\rightarrow (a+x)\phi(a+x), \quad t \rightarrow 0, \quad x > 0; \\ \frac{\partial u}{\partial x} - \left(h + \frac{1}{a}\right) \left[u - \frac{ha}{h + \frac{1}{a}} \psi(t) \right] &\rightarrow 0, \quad x \rightarrow 0, \quad t > 0.\end{aligned}$$

The solution of these equations, obtained from (27), leads to the following value of θ :

$$\begin{aligned}\theta &= \frac{e^{\int_0^t g(\tau) d\tau}}{r(4\pi\kappa)^{1/2}} \left\{ \int_0^t e^{-\int_0^\tau g(\tau) d\tau} \int_a^\infty \rho f(\rho, \tau) \left[\left(e^{-\frac{(r-\rho)^2}{4\kappa(t-\tau)}} + e^{-\frac{(r-2a+\rho)^2}{4\kappa(t-\tau)}} \right) \right. \right. \\ &\quad \left. \left. - 2 \left(h + \frac{1}{a} \right) \int_0^\infty e^{-(h+\frac{1}{a})\xi} e^{-\frac{(r+\rho+\xi-2a)^2}{4\kappa(t-\tau)}} d\xi \right] d\tau d\rho \right. \\ &\quad + \frac{1}{t^{1/2}} \int_a^\infty \rho \phi(\rho) \left[\left(e^{-\frac{(r-\rho)^2}{4\kappa t}} + e^{-\frac{(r-2a+\rho)^2}{4\kappa t}} \right) \right. \\ &\quad \left. \left. - 2 \left(h + \frac{1}{a} \right) \int_0^\infty e^{-(h+\frac{1}{a})\xi} e^{-\frac{(r+\rho+\xi-2a)^2}{4\kappa t}} d\xi \right] d\rho \right. \\ &\quad \left. + ah \int_0^t \frac{e^{-\int_0^\tau g(\tau) d\tau}}{(t-\tau)^{3/2}} \psi(\tau) \int_0^\infty (r+\xi-a) e^{-(h+\frac{1}{a})\xi} e^{-\frac{(r+\xi-a)^2}{4\kappa(t-\tau)}} d\tau d\xi \right\}. \quad (29)\end{aligned}$$

XXXIV. *The Shape of a Liquid Jet.*

By J. M. A. LENIHAN, M.Sc.,
University of Durham (King's College, Newcastle) *.

[Received July 16, 1941.]

ABSTRACT.

A jet of liquid falling at low speed on to a smooth surface assumes a corrugated or beaded shape, hitherto regarded as a stationary wave pattern, caused by progressive disturbances travelling over the surface with the same speed as the falling liquid †. An alternative explanation, based on a graphical solution to Bernoulli's equation, is put forward, and is in agreement with observation.

It is a matter of common observation that an obstacle, such as a stone, on the bed of a stream, gives rise to a pattern of stationary wave type on the surface. The wave-length of the pattern is that corresponding to a progressive train of small-amplitude waves, governed by Kelvin's equation :

$$v = \sqrt{\frac{g\lambda}{2\pi} + \frac{2\pi T}{\rho\lambda}}, \quad \dots \dots \dots (1)$$

in which v = velocity of the stream,

T = surface tension of the medium,

ρ = density of the medium,

λ = wave-length.

An effect similar to this in appearance, but not to be confused with it in theory, is observed when a slowly-moving stream of liquid falls vertically on to a smooth obstacle. The outline of the jet becomes corrugated, giving the appearance of a number of beads hanging one beneath another (fig. 1). This curious property was first examined by Maass ‡, who wrote : " It is based on the phenomenon that an obstacle placed in the path of a jet of liquid flowing vertically downwards causes waves to

* Communicated by Professor George W. Todd, M.A., D.Sc.

† Scott Russel, British Association Report, 1844, p. 362.

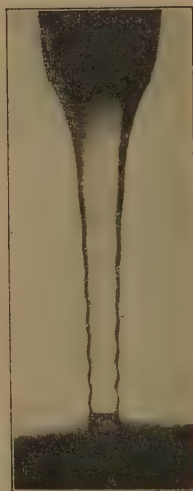
‡ Maass, Roy. Soc. Canada Trans., Ser. III, ix. p. 133 (1915).

travel upwards," thus identifying it with the property of a horizontal stream described above. Maass and subsequent workers * have employed equation (1), and another relation derived by Lamb † for the case of progressive waves on the surface of a cylinder; to obtain values of T from measurements on the jets. Neither of these formulæ can be applied to such cases as that shown in fig. 1, where the length of the "waves" changes considerably between the ends of the jet, and the quantitative investigations so far made have been restricted to long jets which have a number of fairly uniform "waves" near the obstacle (fig. 2). Even here the results obtained are not convincing, and the wave theory is, at

Fig. 1.



Fig. 2.



(NOTE: All the drawings and photographs are reproduced on the same scale—magnified 2.2 times compared with the original jet.)

the best, only a coarse approximation. An alternative explanation will now be attempted.

For any stream of liquid moving without turbulence,

$$p + \frac{1}{2}\rho v^2 + \rho gh = \text{constant}, \quad (2)$$

* Satterly and Macpherson, Roy. Soc. Canada Trans., Ser. III, xxviii. p. 177 (1934); Satterly and Strachan, Roy. Soc. Canada Trans., Ser. III, xxix. p. 105 (1935); Satterly and Strachan, Roy. Soc. Canada Trans., Ser. III, xxx. p. 129 (1936); Prosad Ghosh, Indian Journ. Phys., x. p. 38 (1936).

† Lamb, 'Hydrodynamics,' 2nd Edition, p. 450.

where p =excess pressure across a plane in the stream,

ρ =density of fluid,

u =velocity across the plane in question,

h =depth of this plane below an origin of hydrostatic pressure.

This is Bernoulli's theorem, in which the effects of viscosity are ignored, by assuming the velocity to be constant over any cross-section. If the stream is bounded by curved surfaces, the curvature may be represented by two radii, R_1 and R_2 . In the case of a liquid falling vertically from a circular orifice, the horizontal cross-section is everywhere a circle, and R_1 , R_2 have the same values at all points on the circumference. In these circumstances $p = T \left(\frac{1}{R_1} + \frac{1}{R_2} \right)$, where T is the surface-tension of the liquid used. The radii of curvature may be expressed in terms of differential coefficients in any suitable co-ordinate system, but the equation produced is not capable of general solution. If, as a first approximation, the kinetic energy of the liquid is neglected, the equation (2) above reduces to

$$p + \rho gh = \text{constant.} \quad \dots \dots \dots (3)$$

In this form it is identical with the well-known capillary equation determining the shapes of such bodies as sessile and pendent drops, and may be solved by the following graphical method, due to Kelvin * (see sketch, p. 396).

Construct a vertical line, which will be the axis of symmetry of the figure. Then with centre any point A on either side of the line and radius AB construct a small arc BC. Join CA and let it, produced if necessary, cut the axis at D. Let δz be the vertical distance between B and C, measured downwards. Then evaluate the expression

$$R = \frac{1}{\frac{1}{AB} + \frac{1}{EB} + \frac{\rho g}{T} \delta z - \frac{1}{CD}}$$

and, starting from C, set off this distance CF along CA. With centre F and radius CF draw another small arc CG. GF, produced if necessary, cuts the axis at H. If $\delta z'$ is the vertical distance between C and G the radius of curvature for the next arc is :

$$\frac{1}{\frac{1}{FC} + \frac{1}{CD} + \frac{\rho g}{T} \delta z' - \frac{1}{GH}}$$

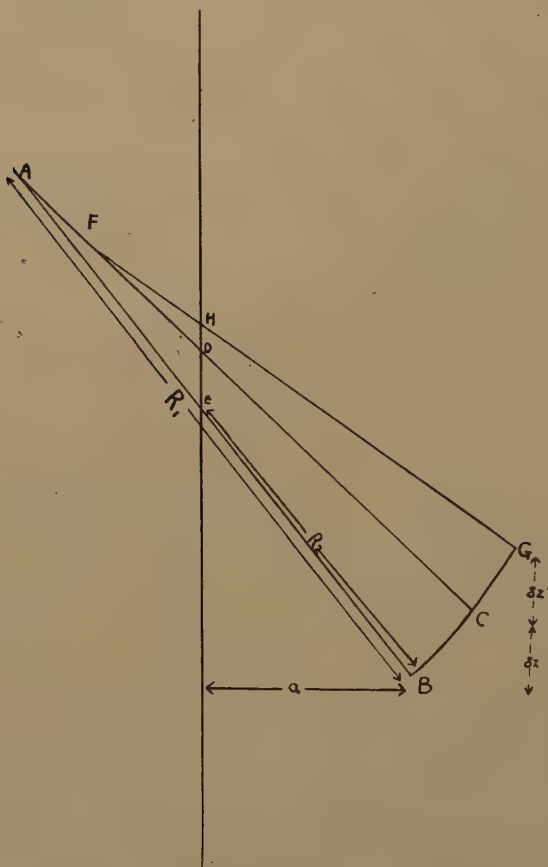
and the centre on the previous radius GF, as before.

* Kelvin, 'Nature,' xxxiv. p. 293 (1886)

This rule may be derived from equation (3) above, for, putting $p = T \left(\frac{1}{R_1} + \frac{1}{R_2} \right)$ and differentiating :

$$\delta \left(\frac{1}{R_1} + \frac{1}{R_2} \right) + \frac{\rho g}{T} \delta h = 0, \quad \dots \dots \dots (4)$$

from which Kelvin's method follows.



Sketch to illustrate Kelvin's method of drawing capillary curves.

Some of the curves constructed by this method are set out in figs. 3 to 6. Fig. 3 corresponds to the observed shape of a drop hanging from a tube, but fig. 6 represents a shape which is highly unstable, on account of the constrictions, and therefore cannot be produced experimentally.

To continue with the investigation of a moving jet, we have, from equation (2) above :

$$\rho gh + T \left(\frac{1}{R_1} + \frac{1}{R_2} \right) + \frac{1}{2} \rho u^2 = \text{constant}, \quad . \quad . \quad . \quad (5)$$

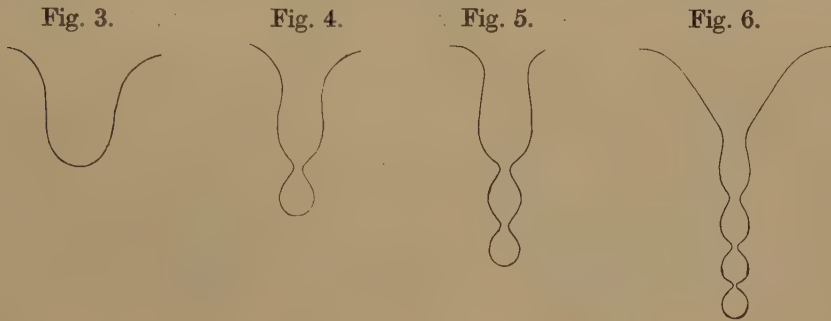
$$\rho gh + T \left(\frac{1}{R_1} + \frac{1}{R_2} \right) + \frac{1}{2} \rho \left(\frac{Q}{\pi a^2} \right)^2 = \text{constant}, \quad . \quad . \quad (6)$$

where Q = volume of liquid delivered per second across a horizontal section of radius a .

Differentiating (6)

$$\frac{\rho g}{T} \delta h + \delta \left(\frac{1}{R_1} + \frac{1}{R_2} \right) + \frac{\rho Q^2}{2\pi^2 T} \cdot \delta \left(\frac{1}{a^4} \right) = 0. \quad . \quad . \quad (7)$$

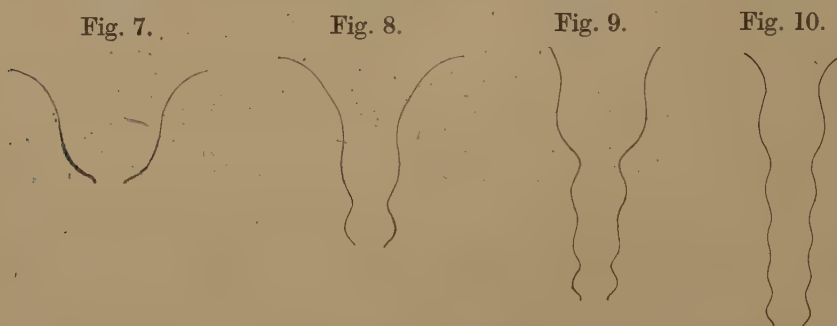
This is similar to equation (4) derived in considering Kelvin's method and leads to a modified rule for the construction of the capillary curve,



in which the term $\frac{\rho g}{T} \delta z$ is replaced by $\frac{\rho g}{T} \delta z + \frac{\rho Q^2}{2\pi^2 T} \cdot \delta \left(\frac{1}{a^4} \right)$. If the volume

of liquid delivered by the jet per second is known, a family of curves may be drawn for various initial values of R_1 , R_2 and a . A difficulty arises over the boundary condition introduced by the obstacle, which is not included in the equations, and a slightly different method of drawing is necessary. In the case of stationary jets the drawings may be commenced at the bottom, where it is known that the figure must close. Here $R_1 = R_2$ and $a = 0$, so that a family of curves may be constructed by varying the initial value of R . For moving jets, where the figure is open at both ends, it is necessary to commence the drawing with arbitrary values for the three variables, R_1 , R_2 and a . The curves may be drawn upwards or downwards from the starting point, and it is found that, as in the case of stationary jets, they open out at the upper end, eventually becoming horizontal and indicating a limit to the drawing. At the lower end, after a certain length has been passed, the curves turn into a

spiral or some other impossible shape. (Impossible, that is, experimentally, though obviously quite sound theoretically.) The drawing must therefore be terminated at the point where this eccentricity commences. Curves constructed by this method are shown in figs. 7 to 10, and it is evident that they bear considerable resemblance to the observed shapes of actual jets. Fig. 11 shows one case in which the resemblance is rather striking; (a) is a drawing of a hanging drop, made by the



Kelvin method, (b) is a drawing of a running jet of the same length, constructed by the method just described, and (c) is a photograph of one of the jets used in this experiment. All are reproduced on the same scale—magnified approximately 2.2 times. The agreement between calculated and observed shapes found by direct measurement in the case of another jet is set out in Table I. Though not exact, the agreement is

Fig. 11.



significant, for the treatment given in this paper is not complete, and the method of drawing makes further inaccuracies unavoidable. It is hoped in experiments now proceeding to obtain further verification of the theory by direct comparison of dimensions.

Apart from this similarity in outline between the photographs and the drawings, there is certain other evidence in support of the explanation put forward.

(1) It will be observed that the equations contain no reference to the diameter of the orifice from which the jet falls, and one drawing should therefore contain the solutions for all permissible diameters. This is the case, for the curves widen out at the upper end, and the shape of a jet falling from a nozzle of given diameter will be the portion of the

TABLE I.

	Total length.	First bead.		Second bead.	
		Length.	Breadth.	Length.	Breadth.
Photograph (fig. 1)	mm. 17.2	mm. 2.02	mm. 2.52	mm. 2.17	mm. 2.63
Corresponding drawing	16.7	1.93	2.67	2.43	2.77

(Both sets of measurements have been reduced to the actual dimensions of the jet; the original drawings were 45 times larger, and the photographs 2.2 times smaller.)

Fig. 12.



Diameter of nozzle=4.22 mm.

Fig. 13.



Diameter of nozzle=10.89 mm.

curve up to the point at which its width becomes equal to that of the orifice. From this it follows that, other things being equal, the shape of a jet should not depend on the size of the nozzle used, except for a small region near the upper end—a conclusion which may be verified by photographing jets from different orifices, with the same rate of flow and the same number of beads. Figs. 12 and 13 are two of a series

obtained in this way, and measurements made on the whole series, set out in Table II., confirm that the shape is almost independent of the nozzle diameter, within the range of sizes used here.

(2) The nature of the obstacle does not enter into any of the equations used to derive the shape of the jet, and it is therefore presumed to have no effect. This conclusion also is in agreement with observation ; jets

TABLE II.

Diameter of nozzle.	First bead.		Second bead.	
	Breadth.	Length.	Breadth.	Length.
cm.	cm.	cm.	cm.	cm.
0.469	0.273	0.211	0.288	0.273
0.654	0.268	0.224	0.283	0.292
0.725	0.269	0.211	0.285	0.294
0.812	0.271	0.226	0.284	0.276
0.900	0.273	0.224	0.289	0.280
1.089	0.271	0.224	0.284	0.284

falling on to different obstacles are identical, except for a small adjustment to the shape of the obstacle at the lower end. In this experiment spherical flasks were used, but any smooth surface allowing rapid drainage of the liquid after impact would be equally suitable.

The author has been greatly assisted by advice and suggestions from Dr. E. G. Richardson and Mr. J. F. Wood.

XXXV. *The Admittance Method for obtaining the Natural Frequencies of Systems.*

By W. J. DUNCAN, D.Sc., A.M.I.Mech.E., F.R.Ae.S.,
Wakefield Professor of Aeronautics at University College, Hull *.

[Received April 22, 1941.]

§ 1. *Introduction.*

THE admittance method of obtaining natural frequencies, which is explained in § 3, is due to Carter and was advanced by him in reference ⁽¹⁾. The method was employed by Carter specifically for the treatment of the oscillations of engine-aircrew systems, and it is now much used in such investigations. However, the method has wider applications, and the object of this paper is to define its field of usefulness. It is shown that it can be used to find the natural frequencies of a system composed of two parts whenever the force system in the connection between the parts depends on a single parameter, for then a unique admittance is definable. When the force system between the parts depends on more than one parameter the simple form of the method fails because a unique admittance no longer exists. However, the method is capable of extension to such cases, but is then considerably more complicated (see §§ 4, 6). A list of some types of system which can be treated by the simple form of the method is given in the Appendix.

The term "admittance" is borrowed from alternating current theory, and in the present context stands for a flexibility measured under the conditions of a steady simple harmonic forced oscillation of the same frequency and phase at all parts of the system considered. Carter himself called this quantity the flexibility, but the latter term is generally understood to apply to static tests, and it seems better to use a distinct term for the quantity appropriate to dynamic conditions. Other writers prefer to use the reciprocal of the admittance, *i. e.*, the dynamic stiffness or impedance. Biot ⁽²⁾ adopts the name "dynamic modulus" for this quantity. It appears that the admittance is slightly more convenient in use than the impedance, and in the opinion of the writer it is undesirable to use a double-worded term such as "dynamic modulus" when an equivalent single word already exists.

* Communicated by the Author.

A point of practical importance is that the admittance characteristics of a system can be found by experiment, or by combined experiment and calculation in the manner of Biot ⁽²⁾. When the characteristics of a pair of systems have been obtained, the natural frequencies of the combination can be predicted.

§ 2. Definition of Admittance.

Consider first, for simplicity and definiteness, a system of flywheels mounted on an elastic shaft. Let a couple proportional to $\sin \omega t$ and of amplitude Q be applied to one end A of the shaft and consider only the steady oscillation of the system having the same frequency as the applied couple. Then if θ is the angular amplitude of the motion at A , the admittance of the system at A is defined to be

$$\gamma_A = \frac{\theta}{Q} \quad \dots \quad (2.1)$$

and is a function of ω . The admittance at any other point is similarly defined in terms of the couple applied at the point and the angular amplitude there.

The definition of admittance can be extended to a general dynamical system as follows:—Let the generalized co-ordinates be $q_1, q_2 \dots q_r \dots$ and the corresponding generalized forces $Q_1, Q_2 \dots Q_r \dots$. Then, if the system moves with a steady oscillation in response to the generalized force $Q_r \sin \omega t$, while all the other generalized forces are zero, the admittance is

$$\gamma_r = \frac{q_r}{Q_r} \quad \dots \quad (2.2)$$

where q_r is now an amplitude.

In all cases the admittance γ is a function of the frequency f (or of $\omega = 2\pi f$), and the relation (or curve) connecting these will be called the admittance characteristic.

§ 3. The Method of Equal and Opposite Admittances.

Suppose that there are two flywheel systems S_1 and S_2 , and that the admittance characteristics of both for end points A_1 and A_2 are known. Then, if the systems are coupled by attaching the shafts rigidly to one another at A_1 and A_2 , any natural frequency of the combined system is such that the admittances of S_1 and S_2 separately for this frequency are equal in magnitude but opposite in sign. For clearly at the point of junction the angular displacements are equal and the torques equal and opposite (Newton's third law). Hence the natural frequencies of the

combined system are found from the points of intersection of the admittance characteristics, one of which must be drawn with the sign of the admittance reversed. A slightly different way of looking at the same fact is as follows:—Take any system of flywheels mounted on an elastic shaft and oscillating in any one of its natural pure modes. Imagine the shaft cut at any point P so as to divide the system into parts S_1 and S_2 . Then the admittances of S_1 and S_2 for the particular natural frequency chosen are necessarily equal and opposite.

§ 4. *Extension of the Method.*

The method just described for finding the natural frequencies of a compound system is not general. It can in fact be applied to many kinds of system, but it will be shown that all of these have the characteristic that the force at the junction is simple, *i. e.*, depends only on a single parameter. An extension of the method can, however, be applied to cases where the force is not simple.

In order to see the limitation of the simple method described in § 3, consider a cantilever beam B_1 performing flexural oscillations, and suppose an extension piece B_2 to be clamped to B_1 at the tip. The first point to be settled is how to define the admittance. Now the action between the beams at the junction has two distinct components, the shearing force and the bending moment. Likewise the displacement at the junction has two distinct components, namely the linear deflexion and the angular deflexion or changes of slope. Thus there are clearly two distinct admittances, namely

$$\gamma_{VP} = \frac{\text{linear deflexion}}{\text{normal force}}$$

and

$$\gamma_{\theta M} = \frac{\text{angular deflexion}}{\text{bending couple}}.$$

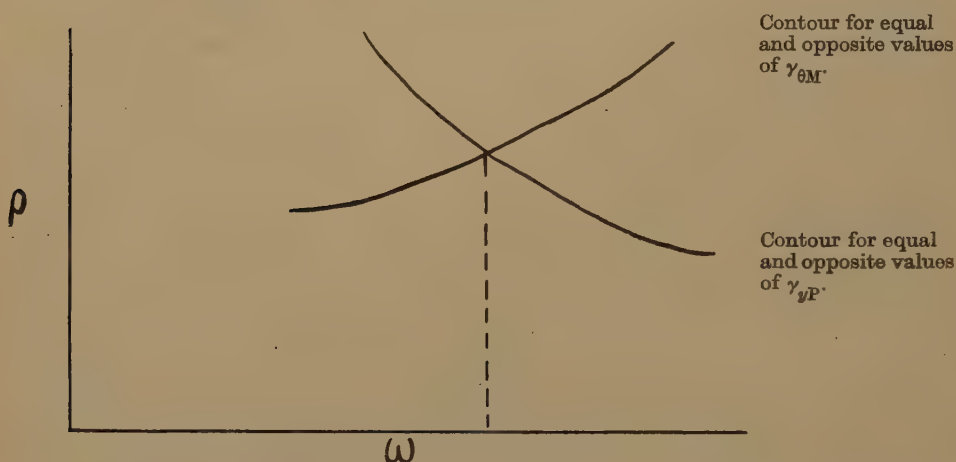
Now it is important to observe that neither of these quantities is definite unless the ratio

$$\rho = \frac{\text{bending couple}}{\text{normal force}}$$

is given as well as the frequency. For displacements of both kinds are generated by the application of either a pure bending couple or a pure normal load. Hence the position is as follows:—Both γ_{VP} and $\gamma_{\theta M}$ are functions of ω and ρ . The corresponding admittances have equal and opposite values for B_1 (cantilever) and for B_2 (free-free beam) when ω and ρ have the values appropriate to any natural pure mode of the combined beams. Thus, in order to find the natural frequencies of the extended beam, it would be necessary to find the two admittances γ_{VP}

and $\gamma_{\theta M}$ for each component as functions of both ρ and ω . The frequencies could be found from the kind of diagram shown in the figure.

A systematic solution can be obtained by the introduction of *cross admittances*. There are in the present problem two displacements (y and θ) and two "forces" (P and M). Take the beam B_1 and apply at the tip unit P and zero M. Then the linear deflexion at the tip is ${}_1\gamma_{yP}$ and the angular deflexion there is ${}_1\gamma_{\theta P}$; the latter coefficient is a cross admittance. The suffix to the left of the symbol indicates the body to which the admittance belongs. When unit M and zero P are applied to B_1 the



linear and angular deflexions are ${}_1\gamma_{yM}$ and ${}_1\gamma_{\theta M}$. By the principle of superposition the deflexions under the action of P and M jointly are

$$\left. \begin{aligned} y_1 &= P {}_1\gamma_{yP} + M {}_1\gamma_{yM}, \\ \theta_1 &= P {}_1\gamma_{\theta P} + M {}_1\gamma_{\theta M}. \end{aligned} \right\} \quad \dots \quad (4.1)$$

Similar formulæ hold good for B_2 , but the signs of P and M must be reversed. Hence the conditions for the equality of the displacements at the junction are

$$\left. \begin{aligned} P({}_1\gamma_{yP} + {}_2\gamma_{yP}) + M({}_1\gamma_{yM} + {}_2\gamma_{yM}) &= 0, \\ P({}_1\gamma_{\theta P} + {}_2\gamma_{\theta P}) + M({}_1\gamma_{\theta M} + {}_2\gamma_{\theta M}) &= 0. \end{aligned} \right\} \quad \dots \quad (4.2)$$

The eliminant of these equations is the frequency equation, for each coefficient γ is a function of the frequency. Clearly the method can be extended to cases where the force system between the bodies depends on any number n of parameters, and the total number of admittances is $2n^2$. It may be noted that so long as all the stiffnesses are due to

conservative forces the admittances for any body will form a symmetrical array.

In the problem discussed above all the forces are applied and all the displacements measured at the same point of the body considered. Also the total number of bodies concerned is two. However, an exactly similar procedure can be followed when the forces are applied to each body at a number of points, and when the system comprises more than two bodies. In such cases it is only necessary to introduce cross admittances which relate a particular component of displacement of one selected point with a particular component of force applied to another selected point. To make this clear, consider a massive and elastic shaft (which need not be uniform) mounted in some definite manner and carrying flywheels of moment of inertia I_1, I_2 at the points P_1, P_2 respectively. Suppose now that a simple harmonic couple of frequency f and unit amplitude is applied at P_1 to the shaft (without the flywheels) and that the angular amplitudes at P_1 and P_2 are γ_{11} and γ_{21} respectively. Similarly the amplitudes at P_1, P_2 due to a couple of unit amplitude and frequency f applied at P_2 are γ_{12} and γ_{22} respectively. If, then, couples of amplitudes C_1, C_2 are applied at P_1, P_2 respectively, the angular amplitudes at these points will be

$$\left. \begin{aligned} \theta_1 &= C_1 \gamma_{11} + C_2 \gamma_{12} \\ \theta_2 &= C_1 \gamma_{21} + C_2 \gamma_{22} \end{aligned} \right\} \quad \dots \dots \dots (4.3)$$

and

Next, suppose that the flywheels are attached to the shaft. Then by consideration of the motion of the flywheel at P_1 ,

$$\begin{aligned} \theta_1 &= (\text{couple applied to flywheel}) \times (\text{admittance of flywheel}) \\ &= (-C_1) \times \left(\frac{-1}{I_1 \omega^2} \right) = \frac{C_1}{I_1 \omega^2}. \end{aligned}$$

Similarly $\theta_2 = \frac{C_2}{I_2 \omega^2}.$

Hence the equations (4.3) become

$$\left. \begin{aligned} C_1 \left(\gamma_{11} - \frac{1}{I_1 \omega^2} \right) + C_2 \gamma_{12} &= 0, \\ C_1 \gamma_{21} + C_2 \left(\gamma_{22} - \frac{1}{I_2 \omega^2} \right) &= 0. \end{aligned} \right\} \quad \dots \dots \dots (4.4)$$

On elimination of C_1 and C_2 the frequency equation of the system is obtained. An exactly similar treatment is applicable to the case where there are n flywheels on the shaft. The frequency equation is then the eliminant of n simultaneous equations similar to (4.4).

§ 5. *Some Examples of the Admittances of Elastic and Massive Bodies.*

The admittances of elastic bodies whose mass is not negligible for the frequencies considered can be obtained by elementary methods in many important cases. A list of some useful admittances is given here.

(a) *Flexural motion of uniform cantilever beam with rigid support at the root.*—Forces applied at the tip and displacements measured there.

Let l = length of beam.

EI = flexural rigidity.

m = mass per unit length.

ω = 2π times frequency.

$$\alpha = \sqrt[4]{\frac{m\omega^2}{EI}} \quad \dots \dots \dots (5.1)$$

$$\gamma_{yP} = \frac{\cosh \alpha l \sin \alpha l - \cos \alpha l \sinh \alpha l}{\alpha^3 EI (1 + \cosh \alpha l \cos \alpha l)} \quad \dots \dots \dots (5.2)$$

$$\gamma_{\theta P} = \gamma_{yM} = \frac{\sinh \alpha l \sin \alpha l}{\alpha^2 EI (1 + \cosh \alpha l \cos \alpha l)} \quad \dots \dots \dots (5.3)$$

$$\gamma_{\theta M} = \frac{\cosh \alpha l \sin \alpha l + \cos \alpha l \sinh \alpha l}{\alpha EI (1 + \cosh \alpha l \cos \alpha l)} \quad \dots \dots \dots (5.4)$$

(b) *Flexural motion of uniform beam with both ends free.*—Forces applied at one end and displacements measured there. With the same notation as for (a).

$$\gamma_{yP} = \frac{\sinh \alpha l \cos \alpha l - \sin \alpha l \cosh \alpha l}{\alpha^3 EI (1 - \cosh \alpha l \cos \alpha l)} \quad \dots \dots \dots (5.5)$$

$$\gamma_{\theta P} = \gamma_{yM} = \frac{-\sinh \alpha l \sin \alpha l}{\alpha^2 EI (1 - \cosh \alpha l \cos \alpha l)} \quad \dots \dots \dots (5.6)$$

$$\gamma_{\theta M} = \frac{-(\cosh \alpha l \sin \alpha l + \cos \alpha l \sinh \alpha l)}{\alpha EI (1 - \cosh \alpha l \cos \alpha l)} \quad \dots \dots \dots (5.7)$$

(c) *Torsional motion of uniform shaft fixed at one end (or connected to an exceedingly massive flywheel).*—Couples applied and displacements measured at two points situated at distances h_1 and h_2 from the fixed end. ($h_2 > h_1$.)

Let l = length of shaft.

J = moment of inertia of unit length.

C = torsional stiffness of unit length.

$$b = \sqrt{\frac{J\omega^2}{C}} \quad \dots \dots \dots (5.8)$$

$$\gamma_{11} = \frac{\sin bh_1 \cos b(l-h_1)}{bC \cos bl} \quad (5.9)$$

$$\gamma_{12} = \gamma_{21} = \frac{\sin bh_1 \cos b(l-h_2)}{bC \cos bl} \quad (5.10)$$

$$\gamma_{22} = \frac{\sin bh_2 \cos b(l-h_2)}{bC \cos bl} \quad (5.11)$$

(d) *Torsional motion of uniform shaft free at both ends.*—Couples applied and displacements measured at two points situated at distances h_1 and h_2 from one end. ($h_2 > h_1$.) With the same notation as for (c).

$$\gamma_{11} = -\frac{\cos bh_1 \cos b(l-h_1)}{bC \sin bl} \quad (5.12)$$

$$\gamma_{12} = \gamma_{21} = -\frac{\cos bh_1 \cos b(l-h_2)}{bC \sin bl} \quad (5.13)$$

$$\gamma_{22} = -\frac{\cos bh_2 \cos b(l-h_2)}{bC \sin bl} \quad (5.14)$$

(e) *Longitudinal motion of a uniform rod fixed at one end.*—Force applied and displacement measured at free end.

Let l = length of rod.
 m = mass per unit length.
 A = area of section.
 E = Young's modulus.
 ρ = density of material.

$$c = \sqrt{\frac{m\omega^2}{AE}} = \sqrt{\frac{\rho\omega^2}{E}} \quad (5.15)$$

Here there is only the single admittance

$$\gamma = \frac{1}{EA c} \tan cl. \quad (5.16)$$

(f) *Longitudinal motion of a uniform rod free at both ends.*—Force applied and displacement measured at one end. With the same notation as for (e).

$$\gamma = -\frac{1}{EA c} \cot cl. \quad (5.17)$$

§ 6. Some Examples of the Use of Admittances.

(i) *Rigid body attached to tip of uniform cantilever beam.*—It will be assumed that the body is rigidly attached to the beam at the centre of mass of the body, so that the latter partakes of the angular as well as

of the linear motion of the beam-tip. It may be noted that the analysis remains unaltered if the point of attachment of the body is not the centre of mass, provided that the line joining it to the centre of mass is perpendicular to the beam.

The admittances of the beam (which will be written with the suffix 1) are as given in § 5, Example (a). The admittances of the body are clearly

$${}_2\gamma_{\theta P} = -\frac{1}{\omega^2 M}, \quad {}_2\gamma_{\theta M} = -\frac{1}{\omega^2 J},$$

$${}_2\gamma_{\theta M} = {}_2\gamma_{\theta P} = 0,$$

where M is the mass of the body and J the moment of inertia about the point of attachment. If the centre of mass is offset from the tip of the beam, and the line joining it to the tip is not perpendicular to the beam, then the admittances ${}_2\gamma_{\theta M}$ and ${}_2\gamma_{\theta P}$ will not vanish. This case will not be pursued, but offers no difficulty.

When the values of the admittances are substituted in equations (4.2) and P and M eliminated, the result obtained after reduction, namely the frequency equation of the system, is

$$\begin{aligned} & (1 + \cosh \beta \cos \beta)^2 \\ & - \left(\frac{\beta^3 J}{l^3 m} \right) (1 + \cosh \beta \cos \beta) (\cosh \beta \sin \beta + \cos \beta \sinh \beta) \\ & - \left(\frac{\beta M}{lm} \right) (1 + \cosh \beta \cos \beta) (\cosh \beta \sin \beta - \cos \beta \sinh \beta) \\ & + \left(\frac{\omega^2 M J}{m E I} \right) (\sin^2 \beta - \cos^2 \beta \sinh^2 \beta) = 0, \quad \dots \dots \dots (6.1) \end{aligned}$$

$$\text{where} \quad \beta = \alpha l = \sqrt[4]{\frac{m \omega^2 l^4}{E I}} \quad \dots \dots \dots (6.2)$$

β and all of the coefficients in (6.1) are non-dimensional.

(ii) *Pair of flywheels attached to a uniform shaft which is fixed at one end.*—The problem has already been discussed in a general way at the end of § 4. When the values of the admittances given in example (c) of § 5 are substituted in equations (4.4) and C_1 and C_2 eliminated, the result reduces to

$$\begin{aligned} \cos bl - \left(\frac{b I_1}{J} \right) \sin b h_1 \cos b(l-h_1) - \left(\frac{b I_2}{J} \right) \sin b h_2 \cos b(l-h_2) \\ + \frac{b^2 I_1 I_2}{J^2} \sin b h_1 \cos b(l-h_2) \sin b(h_2-h_1) = 0. \quad (6.3) \end{aligned}$$

(iii) A uniform rod is fixed at one end and another uniform rod of different section, etc., is attached at the free end.—Let the suffix 1 apply to the rod with the fixed end, and the suffix 2 to the other. Then

$$\gamma_2 = -\gamma_1,$$

and on substitution of the values of the admittances given in examples (e) and (f) of § 5 this becomes

$$\left. \begin{aligned} \tan c_1 l_1 \tan c_2 l_2 &= \sqrt{\frac{m_1 E_1 A_1}{m_2 E_2 A_2}}, \\ &= \frac{A_1}{A_2} \sqrt{\frac{\rho_1 E_1}{\rho_2 E_2}}. \end{aligned} \right\} \quad (6.4)$$

APPENDIX.

Some Types of System which can be treated by the Simple Form of the Method.

As explained in § 4, the necessary condition is that the force system in the connexion between the bodies shall depend on a single parameter. A few such cases are listed and the nature of the force is indicated in parentheses :—

Flywheels on a shaft. (Torsional couple.)

System of loaded or unloaded rods of constant or variable section performing longitudinal oscillations. (Longitudinal force.)

Transverse oscillation of a loaded or unloaded cord of constant or variable section under tension. (Transverse component of force.)

Oscillation under gravity of a multiple pendulum. (Horizontal component of force.)

Beam attached symmetrically and at right-angles to another beam, *e. g.*, aeroplane tail on fuselage. (Normal force.)

Numerous electrical problems can also be treated by the method.

References.

- (1) Carter, "The Vibration of Airscrew Blades with particular reference to their Response to Harmonic Torque Impulses in the Drive." Reports and Memoranda of the Aeronautical Research Committee, No. 1758, July, 1936.
- (2) Biot, "Coupled Oscillations of Aircraft Engine-Propeller Systems." Journ. of Aero. Sci. vii. No. 9 (July, 1940).

XXXVI. *On a Method of Correcting the Spherical Error of Electron Lenses, especially of those Employed with Electron Microscopes.*

By R. KOMPNER *.

[Received August 8, 1941.]

Introduction.

IT seems as if the electron microscope, constructed on more or less conventional lines, will soon have reached the limit of its resolution. This limit appears to depend on the inevitable spherical error of all electron lenses and on electron diffraction on the one hand, and on chromatism (of the irradiating electron beam, or caused by dispersion when travelling through the object) on the other hand. Various other methods, such as the "shadow" microscope or the "probe" microscope, are subject to the same limitations and do not affect the purpose of this study.

Leaving chromatism aside for the time being, we note that spherical error and diffraction tend to act in opposite directions. That is to say: an increase in the size of aperture will cause an increase of the spherical error, and a decrease of the error due to diffraction, and *vice versa*; hence, for a given electron velocity and type of objective lens there is one, and only one, size of aperture for optimum resolution. With this aperture the error due to spherical aberration and the error due to diffraction are equal. As objectives having less spherical error are being developed, the aperture subtended by the irradiating electrons can be increased, and thus some progress towards better resolution be made.

If it were possible to extend the resolution obtainable at present ($\approx 5 \times 10^{-7}$ cm.) by about one order of magnitude, an important advance would have been made. Such a resolution would make possible the investigation of objects of *atomic* dimensions, a most desirable goal.

The subsequent suggestions are made in the hope of contributing to the attainment of that goal.

Brief Description of Idea.

The principal idea is, in short, to illuminate the object with electrons of various velocities: the velocities are to increase with the distance from the axis, when passing through the lens, in such a way as to compen-

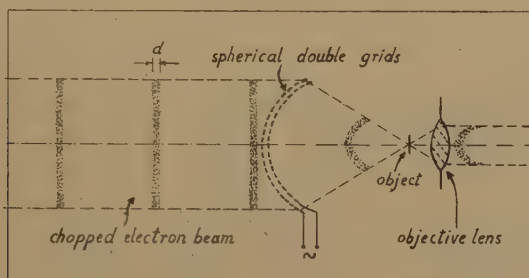
* Communicated by the Author.

sate exactly for the apparent increase above linearity of refracting power of the lens (which is the cause of spherical error).

It is perhaps impossible to achieve such a velocity distribution by purely "static" means. But a solution of this problem will be proposed which makes use of "dynamic" means, *i. e.*, of the so-called phase-focusing or velocity modulation. It will be shown that one form of chromatism, namely that of the irradiating beam, can be corrected at the same time *.

Let us suppose a parallel beam of electrons of uniform velocity to be "chopped," *i. e.*, interrupted into flat disc-like electron clouds and empty spaces, see fig. 1. When these plane clouds impinge on spherical double grids, which are supplied with alternating voltage of very high frequency (the so-called modulating voltage), successive annular portions of the clouds will be accelerated or slowed down differentially, according

Fig. 1.



to the actual potential difference between the grids at any instant of time. To a first approximation, the velocities of the electrons resulting from the grid potentials will depend on the square of the original distance from the axis.

This spherical double grid will concentrate the electrons into some kind of focal point, at which the object is placed. After that, the electrons will diverge again to pass through the objective lens, which will now be traversed by electrons having velocities to compensate its non-linearity or spherical error at any point.

To be able to judge of the practicability of the method proposed above, it will be necessary to go into more detail, such as the actual order of magnitude of the lens distortion causing the spherical error.

* A method has indeed been suggested, which eliminates chromatism generally by using an electro-static objective lens having superposed on its initial lens potential difference a sinusoidal voltage of ultra high frequency, in a paper on "Achromatic Electrons Lenses," by A. Nesslinger, *Jahrb. d. A. E. G. Forsch.* vi. p. 83 (1939).

Magnitude of Spherical Error.

We find fairly universal agreement in literature concerning the electron microscope that the resolution δ , *i. e.*, the dimensions of the smallest discernible detail in the object in cm. attainable in consequence of the spherical error of a lens, can be expressed by a law

$$\delta = C \cdot \alpha_0^3, \quad (1)$$

where C is a constant depending on the type of lens used and on the position of the object in relation to the lens, which constant is determined by experiment in every case, and α_0 the angle or aperture subtended by the peripheral electrons and the axis.

The longitudinal aberration, due to the spherical error denoted by Δf , is said to depend on the radius of the zone of the lens in action, and is given by

$$\Delta f / f = C' \cdot r^2 = C' \cdot f^2 \cdot \alpha_0^2. \quad (2)$$

For geometrical reasons, we have $\Delta f / f \approx \delta / r$.

Hence
$$C' = \frac{C}{f^3} \quad (3)$$

Bories and Ruska * give values for C between 1500 and 15; allowing for the fact that they calculate the resolution in mm. we may safely assume C to be about 10, with a focal length, which should be easily attainable, of $f = 0.5$ cm. We thus obtain

$$C' \approx 10 / 0.5^3 = 80.$$

In this way we are able to estimate the deviation of the refracting power of the lens from linearity with increasing lens radius by observing the longitudinal aberration. Remembering that with magnetic lenses the focal length is directly proportional to the potential difference causing the electron velocity, we can write for the necessary additional potential difference ΔU_0 to correct spherical error

$$\frac{\Delta U_0}{U_0} = C' \cdot f^2 \cdot \alpha_0^2 = 20 \alpha_0^2. \quad (4)$$

The wave-length λ of the electron beam is given by the De Broglie equation

$$\lambda = \frac{1.23}{\sqrt{U_0}} \cdot 10^{-7} \text{ cm.}, \quad (5)$$

and the resolution—if diffraction only were active—to be expected, according to Abbe, is

$$\delta \approx \frac{0.5 \lambda}{\alpha_0} \text{ cm.} \quad (6)$$

* "Resolving Power of the Super Microscope," *Zeit. Schr. f. Techn. Phys.* p. 227 (1939).

Let us assume U_0 to be $\approx 4 \cdot 10^4$ volt; this gives $\lambda \approx 6 \cdot 10^{-10}$ cm., and taking $\alpha_0 = 1/50$ then $\delta = 1.5 \times 10^{-8}$ cm. *

To correct the lens completely for spherical error up to an aperture of 1/50th, the electrons will have to be accelerated by potentials depending on the square of the aperture, from U_0 at the axis up to $U_0 + \Delta U_0$ at the periphery, where

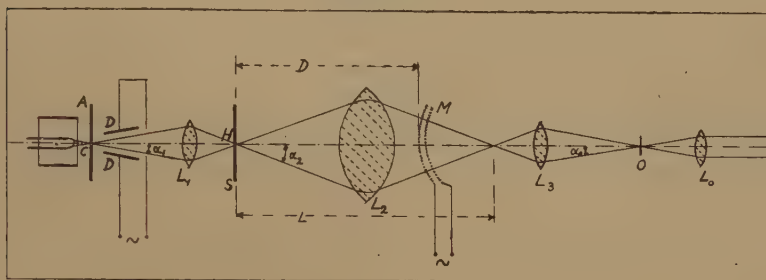
$$\frac{\Delta U_0}{U_0} = \frac{20}{50^2} = 8 \cdot 10^{-3}.$$

That means : $\Delta U_0 \approx 320$ volts.

To arrive at the frequency of the modulating voltage we have to consider the actual physical dimensions of the apparatus.

An arrangement slightly different from the one proposed before, although basically the same, is shown in fig. 2. An electron beam is

Fig. 2.



produced in the conventional way by a thermionic cathode and accelerated and focused into a so-called "cross-over" C by the anode A. The cross-over is imaged upon a shield S, by a lens L_1 , to coincide directly with a small hole H bored into the shield S. Deflecting plates D—D are supplied with an alternating potential which causes the image of the cross-over to sweep to and fro across the shield, allowing the electrons

* It will have to be left to experiment to decide whether there will be an error due to diffraction at all, and what its magnitude will be. As every point of the object and the image is traversed by electrons of various velocities and wave-lengths, it can be reasoned that these electrons will fail to interfere in the formation of the image. On the other hand we have, at any instant, only electrons of some particular uniform velocity passing through the object and at some later instant forming the image. Thus it can be expected that there will be an error due to diffraction which, however, will vary periodically with time, as electrons with differing wave-lengths succeed each other. As the velocity variation is small compared with the unvaried velocity, the order of magnitude of error due to diffraction is assumed to be given by the unmodulated voltage U_0 .

There will have to be more than two stages of magnification, most likely three stages, each magnifying about 100 times, giving an electron-optical magnification of about 10^6 followed by a light-optical stage of 10, altogether with a total magnification of 10^7 .

Correction of Chromatism.

The chromatism due to the initial thermionic velocity distribution ΔV_0 can be corrected if ΔV_0 is much smaller than V_0 . With thermionic cathodes $\Delta V_0/V_0$ is usually of the order of 10^{-2} and less.

Let us assume that we succeed in producing practically infinitely thin electron clouds by the methods for "chopping" proposed above. Then the thickness d of a cloud arriving at the modulating grid is given by

$$d = \frac{\nabla V_0}{V_0} \cdot D, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

where D is the distance from the origin of "chopping" to the modulating grids. Let us assume, moreover, that the modulating grids are very near together.

The cloud will consist of layers of different velocities; the hindmost layers being the slowest ones will be accelerated most, since the potential difference between the grids will have increased during the time $T = \Delta V_0 / D$ of passage through the grids. Therefore, the electrons emerging at *any point* of the modulating grids will be isochromatic, *i. e.*, of uniform velocity, if the following condition is satisfied :

$$D = 2\tau_0 V_0 / \frac{\Delta U_0}{U_0} = \frac{L}{40} (\alpha_2/\alpha_0)^2. \quad (9)$$

With the values assumed we obtain $D=250$ cm. ; this value is too large although of the right order of magnitude. A slight alteration, for instance, in the factor of the spherical error estimated originally as $C'=80$, may bring D down to a reasonable dimension. It will be observed that the correction applies *irrespective and independent of the actual value of ΔV_0* for small $\Delta V_0/V_0$.

Chromatism caused during the passage of the electrons through the object will *not* be corrected by this method. Completely opaque objects, whose silhouettes will be made visible, are therefore most suitable. It may be found, however, that electrons dispersed and slowed down by the object are distributed over larger areas of the final image, thus merely reducing its contrast.

"Chopping" and Intensity Questions.

Since the basic assumption was made, namely, that it is possible to produce electron clouds of practically infinitely thin thickness, it remains to be investigated how far this assumption can be justified.

The image of the cross-over, having a diameter of less than 10^{-2} cm., is made to sweep across the shield S with a frequency $f=2.4 \times 10^8$ cycles. If we assume a sweep amplitude of $A=1$ cm., the time of transit across a hole of a diameter of $H=10^{-2}$ cm. is given as

$$t_0=10^{-2}/(\pi \times 2.4 \times 10^8)=1.33 \times 10^{-12} \text{ sec.}$$

or the initial thickness

$$d=t_0 \cdot V_0 = \frac{H \cdot V_0}{\pi A f} = 1.6 \times 10^{-2} \text{ cm.} \quad . \quad . \quad . \quad (10)$$

Since the final thickness d is of the order of $10^{-2} \times D$, *i. e.*, about 1 cm., the assumption is quite justifiable.

The calculations made above prompt at once the question: will there still be sufficient current to give a picture observable on fluorescent screens?

The beam current integrated over time, it is true, will be reduced by a factor

$$t_0 \times f = \frac{10^{-2}}{\pi} \approx 3.2 \times 10^{-3},$$

but it has to be remembered that the aperture of the total instantaneous illuminating beam has been increased by a factor of about 20 over the apertures used at present. That means that the instantaneous currents may be increased by a factor of about 4×10^2 , which gives us an illuminating intensity quite comparable with intensities obtained at present.

Final Remarks.

Several difficulties have not been discussed—so far—such as: firstly, the directional dispersion which will be caused by grids as proposed; secondly, the spherical error caused by the lenses L_1 , L_2 , L_3 ; thirdly, the effect of the electrostatic repulsion of electrons upon each other; fourthly, the error caused by approximating the cosine function by the square function. But it is felt that it would be premature to go too far into details of a suggestion which is somewhat utopian at the present time.

It is hoped, however, that it has been shown that there is, at least theoretically, the possibility of extending the range of electron microscopy.

XXXVII. *A Theory of the Anode Fall in Glow Discharges.*

By A. VON ENGEL *.

[Received August 22, 1941.]

1. *Introduction.*

THIS paper attempts to work out the theory of the anode fall † in which attention is restricted to the main processes occurring in the neighbourhood of the anode when a positive column is present, *i. e.*, when the anode is so remote from the cathode that it lies beyond the Faraday dark space. In particular, an attempt is made to find the factors that govern the dependence of the anode fall of potential on the current between the electrodes (the so-called "characteristic" of the anode fall), taken as usual to denote the potential difference between the anode and the boundary of the positive column.

For convenience of description, the whole region of the discharge between the anode and the not very sharply defined positive boundary of the positive column is termed the "anode zone": the boundary between the positive column and the anode zone can then be defined as the region in which the electric field ceases to have the constancy characteristic of the positive column. The visual appearance of the anode zone varies: sometimes a dark or relatively dark space exists on the side of the positive column which merges into the anode glow, the radiation from which is due to the acceleration of electrons towards the anode; and it is possible that in some cases a narrow dark space may exist between the anode glow and the anode which is somewhat similar to Aston's dark space at the cathode ‡. The anode zone is to be regarded as a region similar to that set up round a positive probe introduced into a plasma.

A simplified representation is that of a discharge maintained between plane electrodes at the ends of a cylindrical (insulating) tube by constant

* Communicated by Professor Sir J. S. Townsend.

† Up to the present no complete theory concerning the anode region exists; see for example Druyvesteyn and Penning, ref. ⁽⁸⁾, as well as refs. ^(1-4a). The basic ideas of the following investigation originate during the later years of my collaboration with M. Steenbeck.

‡ Concentration of the current in spots on the anode will not be considered here, see refs. ^(5,6).

potential difference, the gas being at a pressure of a few tenths of a mm. of mercury or less, and the current density 10^{-3} amperes per cm.² or less. The positive column is supposed to fill the tube except for the wall-sheath, to be uniform so that the longitudinal electric field is constant, and to be free from oscillations and striations. The gas is supposed to have no electron affinity, so that the conductivity is due exclusively to the motion of (positive) ions and electrons. As in the case usually realized by experiment, the region of highest (positive) potential is that of the anode itself.

In the anode zone electrons are accelerated towards, and ions away from, the anode. The migration of ions towards the positive column is maintained by ions generated round the anode by collisions between the accelerated electrons and molecules of the gas. But only a small number of ions is necessary compared with that of the electrons which pass the cross-section of this zone per unit time. On the other hand, the number of electrons and ions in the positive column per cubic centimeter is large and practically equal, whereby the concentration decreases towards the anode due to the higher speed of the particles there. It seems as if these processes set up a local concentration of electrons close to the anode surface and a positive ion cloud increasing in density at a progressively diminishing rate towards the negative boundary of the anode zone. In consequence of this disturbance of the total and individual space charges in the anode zone, a varying electric field is set up, the line integral of which across the zone is termed the anode fall. This field and the resulting potential differences cause the further generation of ions and electrons by collision; and since the electrons required for the transport of the current are relatively small in number, this source of negatively charged particles may be regarded as relatively indefinitely large, a concept which will be seen later to lead to a great simplification of the present problem. Additional simplification is gained by ignoring a number of secondary processes such as recombination, wall effects, excitation, collisions of the second kind, diffusion of ions and electrons between the anode zone and the positive column, the liberation of gas from the anode, and liberation of electrons from the anode by impact of electrons.

2. Theory of the Anode Fall.

According to our representation, the field strength in the positive column is small compared with the mean field strength in the anode zone. We assume, corresponding to the findings of Güntherschulze, Bär, and Betz *, that E_x , the field strength in the anode zone, is parabolic and depends on one coordinate x only. Taking the origin in the plane of

* Ref. (5), p. 298 ff.

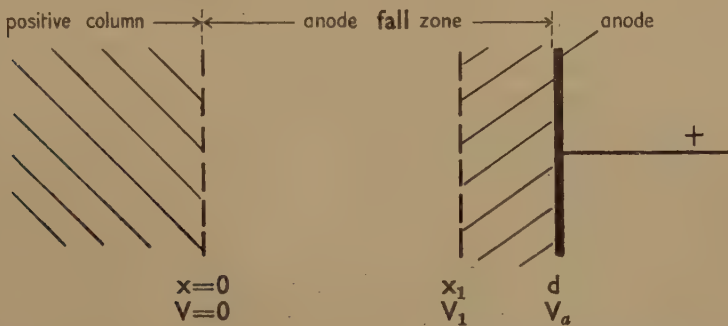
the positive column, the boundary of the anode zone (see figure), we then have at distance x , measured towards the anode,

$$E_x = A \left[\left(\frac{d}{2} \right)^2 - \left(x - \frac{d}{2} \right)^2 \right] = \frac{3}{2} \frac{V_a}{d} \left[1 - \left(\frac{2x}{d} - 1 \right)^2 \right], \quad (1)$$

where V_a is the potential of the anode, V_x is taken to be zero at $x=0$, d is the value of x at the anode, and $A=1.5 V_a/d$. V_x is then :

$$V_x = 3V_a \left(\frac{x}{d} \right)^2 \left[1 - \frac{2x}{3d} \right] \quad (2)$$

It is next assumed that the rate of generation of new charged particles by electron impact may be calculated from the data for ionization provided by electron beam experiments. To a first approximation, the number



of ionizing collisions (dN) occurring in a distance dx at the electron speed of V electron volts is given by

$$dN = ap(V - V_i)dx, \quad (3)$$

where p is the gas pressure in mm. of mercury, a a constant peculiar to the gas, and V_i the ionization potential in electron volts. In applying this formula to the present problem it is assumed that the mean energy of the electrons crossing the boundary $x=0$ of the anode zone is V_s volts*, equal to the mean kinetic energy of the electrons in the positive column. It then follows that at any point in the anode zone we may write

$$dN = ap(V_x + V_s - V_i)dx. \quad (3.1)$$

From (3.1) we may derive the interesting quantity N , the total number of ionizing collisions (or the total number of new ions and electrons generated per electron) occurring between some plane $x=x_1$, at which

* The exact calculation, using the Maxwellian distribution, leads to the same result.

such collisions first become significant, and the anode. In calculating x_1 we suppose that in the whole of the anode zone electrons move mainly in the direction of the field because this is so large; in addition, allowance should be made for the fact that the term V_s in (3.1) is associated with the random motion of the electrons in the positive column. When x_1 is slightly less than d , which is the case to be expected in experiment, we may write

$$x_1 = d(1 - \epsilon), \quad (4)$$

where $\epsilon \ll 1$. It follows that V_1 , the potential at $x = x_1$ is given by

$$V_1 = V_i - V_s,$$

and hence it follows from (2) that

$$V_1 = 3V_a \left(\frac{x_1}{d} \right)^2 \left(1 - \frac{2x_1}{3d} \right) (4.1)$$

From (4), after some re-arrangement, and neglecting terms of the second order, it is then found that

$$\epsilon = [(V_a + V_s - V_i)/3V_a]^{1/2}. \quad (5)$$

On integrating (3.1) between the limits x_1 and d , where x_1 is derived from (4) and (5), we find

$$N = apd(V_a + V_s - V_i)^{3/2}/(3V_a)^{1/2}. \quad (6)$$

It is here supposed that there is no loss of ions and electrons, that is, that recombination in the gas and wall effects may be neglected. Provided that the concentration of positive ions near the anode is large, the régime in the anode zone is then comparable with the case (considered first by Langmuir) of two plane electrodes opposite each other in a gas at very low pressure, one electrode being an indefinitely copious source of electrons and the other of positive ions. For this case (*cf.* ref. (2)) the ratio j_+/j_- of the ionic and electronic current densities is given by

$$j_+/j_- = (m_-/m_+)^{1/2}, \quad (7)$$

where m_- and m_+ denote the masses of the two kinds of carrier. It follows that for many gases j_+ is of the order of 1 per cent. of j_- , so that j , the total current density, differs very little from j_- ; and this, we conclude, represents also the conditions to which the present model relates.

For the same model it has been shown * that

$$j_- = \frac{3, 5}{16\pi} \left(\frac{2e}{m_-} \right)^{1/2} \frac{V^{3/2}}{d^2}, \quad (8)$$

* The numerical constant in (8) is different from that in equation (208) of v. Engel-Steenbeck, 'Elektrische Gasentladungen,' vol. 1. The recalculation of

$$\int_0^U dV/[V^{1/2} + (U - V)^{1/2} - U^{1/2}]^{1/2}$$

where V is the electrode potential difference and d is now the electrode spacing. We now suppose that we may apply (8) to the anode zone at least to a first approximation, so that V becomes V_a and d becomes the distance between two virtual electrodes.

Since N is the total number of positive ions generated by one electron when passing the anode zone, it follows that in a steady state N also represents the ratio of the number of the positive ions per cm^2 per sec. to the number of electrons per cm^2 per sec. in any cross-section between the two virtual electrodes. From that, and from equation (6) and (7), we have

$$N = (m_-/m_+)^{1/2} = apd(V_a + V_s - V_i)^{3/2}/(3V_a)^{1/2}. \quad (9)$$

On substituting the expression for d given by the above relation in (8), and neglecting the difference between j and j_- , we find

$$j/p^2 = \frac{1}{16\pi} \frac{m_+}{m_-} \left(\frac{2e}{m_-}\right)^{1/2} a^2 V_a^{1/2} (V_a + V_s - V_i)^3, \quad (10)$$

which is the required relation giving the dependence of V_a on j . Since V_a is often little different from V_i , the above expression may be written in the simple approximate form

$$j/p^2 = \text{const. } V_i^{1/2} (V_a + V_s - V_i)^3. \quad (11)$$

The dependence of the thickness of the anode zone on the anode fall follows directly from (9). The influence of the current density on the anode zone is shown by the following equation, which results from (8) :

$$pd = \left(\frac{3}{16\pi}\right)^{1/2} \left(\frac{2e}{m_-}\right)^{1/4} V_i^{3/4} \left(\frac{1}{j/p^2}\right)^{1/2}, \quad (12)$$

in which $V_a^{3/4}$ is replaced by $V_i^{3/4}$. It may be recalled that if electrostatic units are used for charge, current and tension, the thickness is expressed in cm.

3. Discussion.

The equations (11) and (12), expressed as usual in terms of the rule of similarity, show :—

(a) The anode fall of potential is not very different from the ionization potential ; it rises slowly and steadily with increasing current density and decreases with the gas pressure ; it is not dependent on the anode material.

has been carried out recently by Simpson's rule. The result of the calculation is as follows :—The space charge limited electron current between two plane electrodes at low pressure (plate distance $\ll$ mean free path of electrons or ions) is approximately doubled (and not increased five times) when the anode plate acts as a source of positive ions.

(b) The anode fall characteristic $V_A=f(j)$ does not show an effect analogous to the normal cathode fall.

(c) The anode fall depends on the random electron velocities in the positive column.

(d) The thickness d of the anode zone and the current density j are related by the equation

$$d \times j^{1/2} = \text{const.},$$

so that the thickness is independent on the gas pressure.

As an example the anode fall characteristic of He will be determined ; by introducing the constants

$$\text{Ionization potential} \quad V_i = 24.5 \text{ V}$$

$$m_+/m_- = 7.28 \times 10^3$$

$$a = 4.6 \times 10^{-2}$$

pairs of ions per cm., volt and mm. Hg. (see ref. (2), v. Engel-Steenbeck, vol. 1, p. 38), and assuming a discharge tube with a radius of a few cm. and a gas pressure of 0.5–1 mm. of mercury ($V_s \approx 4.5$ electron volts*), we have, from (10) :

$$V_A = 20 + 20.5 \left(\frac{j}{p^2} \right)^{1/3}.$$

It follows from the last equation that the anode fall at 1 mA/cm.² and 1 mm. Hg is only 2 V higher than at a very small current density.

The thickness of the anode zone for He is according to (12), when j is expressed in A/cm.².

$$d = 2.4 \times 10^{-2} \left(\frac{1}{j} \right)^{1/2}.$$

At 1 mA/cm.², d becomes 0.76 cm.

There exist up to the present only a few experimental verifications of the theoretical results. The anode fall of potential is determined by measuring the steep fall of voltage which appears when the anode gradually penetrates into the Faraday dark space, or by measuring the potential in the neighbourhood of the anode with probes, or by calorimetric measurement of the heat absorbed by the anode. The thickness of the fall zone is measured visually or with probes.

The table on p. 423 gives some experimental results which should be compared with the calculated ones. From them it is evident that the anode fall is always just a little higher than the ionization potential of the gas, if the current density and the pressure are not too high. The independence of the thickness of the anode zone and the pressure seems

* V_s can be calculated by means of the electron temperature function given in ref. (2), vol. 2, p. 86.

Gas.	p mm. of mero.	j A/cm. ²	V_a v.	V_i v.	d cm.		Reference.
					Measured.	Calculated.	
Hydrogen ..	0.2	3×10^{-3}	≈ 25	$\begin{cases} H_2 : 15.5 \\ H_1 : 13.5 \end{cases}$	—	0.2	Bramley (11)
	1-16	$\approx 10^{-4}$	16.7		—	≈ 1.1	Güntherschulze (5a)
	1-5	$\approx 10^{-4}$	17-19		0.5-0.7	≈ 1.1	Güntherschulze (5)
Deuterium..	1-5	$\approx 10^{-4}$	16-18	Probably the same as V_i of hydrogen.	0.6	1	Güntherschulze (5)
Nitrogen ...	0.8	4×10^{-4}	16.5	$\begin{cases} N_2 : 15.8 \\ N_1 : 14.5 \end{cases}$	—	0.9	Güntherschulze (5a)
	0.5	—	15.7		—	—	Rubens (6); anode spot !
Oxygen	1 order 0.1 (?)	6×10^{-4} $\approx 10^{-4}$	14.2 —	$\begin{cases} O_2 : 12.5 \\ O_1 : 13.5 \end{cases}$	— > 0.12	0.4 1.	Güntherschulze (5a) Aston, <i>l. c.</i>
Helium ...	10	1.6×10^{-3}	26	24.5	—	0.5	Druyvesteyn (7)
Neon	10	8×10^{-4}	17	21.5	—	0.5	Druyvesteyn (7)
Argon.....	20 0.5	1.6×10^{-3} 1.2×10^{-3}	10 15.3	15.7	— —	0.43 —	Druyvesteyn (7) Partzsch (13)

REMARKS.—Measurements in mercury vapour uncertain [V_a between 13 and 3 V; Matthies, Langmuir (4a), Partzsch (13)]. When comparing V_a and V_i dissociation has to be considered. d calculated from equation (12).

to be confirmed by Aston's * observation on the glow discharge between aged aluminium electrodes in oxygen. He observed between the anode and the positive glow near the anode a dark space with a sharp edge when the anode penetrated the bright negative glow. Under these conditions the thickness of the anode dark space was independent of the pressure, the range of which has not been given (presumably 0.1 to 1 mm. pressure). At 0.1 mA/cm.² the thickness was $d=1.2$ mm. in pure oxygen, and he estimates that it varies inversely as the square of the current density, contrary to (12). The difference between this observation and our theory may be due either to the effect of negative ions, or to the fact that (12) applies to the thickness of the whole fall space and not to that of the dark space, or to the penetration of the anode into the Faraday dark space. In air and nitrogen less definite dark spaces were observed, but that was not the case in hydrogen.

Although Aston did not actually make observations on the thickness of the whole anode zone, but only of a part of it, the analogy between our considerations and his observations is obvious. As mentioned in the introduction, the anode zone often consists of a small dark space and an extensive anode glow. It is noteworthy that the observed thickness of the dark space is much smaller than the expected thickness of the anode zone. Aston's observation on the sharpness of the edge of the dark space after a few months use might be explained by supposing that after this period the aluminium anode became oxidized by chemical reactions (ref. ⁽¹²⁾) between the gas and the metal anode, whereas on fresh electrodes reaction might cause emission of light concealing the visual appearance of the anode dark space.

In recent years Güntherschulze and Schnitger † carried out measurements on the anode fall and the electrical and visual thickness of the anode fall zone in hydrogen and deuterium, in a discharge without a positive column between iron electrodes in a tube with a radius of 12.5 cm. The anode fall is in both cases just a little higher than the ionization potential of H, and the thickness is about half of the theoretically expected one, results which fit in well with the suggested theory. The fact that the anode fall in deuterium is found about 1 volt less than in hydrogen at pressures within 1 and 5 mm. Hg might be due to the greater value of m_+ in (10) ‡. All these experiments were made at a current

* F. W. Aston, 'Nature,' lxxxix. p. 218 (1912); see also ref. ^(5a): it has been found that the thickness of the anode glow is independent of the gas pressure.

† A. Güntherschulze and H. Schnitger, *Zeits. f. Phys. civ.* p. 395 (1936).

‡ Assuming that V_i , V_a and a are numerically equal for hydrogen and deuterium, and allowing for the doubling of the mass of the ion, it follows from (10) that V_a in deuterium is 1 to 1.3 volts smaller than in hydrogen.

density equal to that at the normal cathode fall [about 10^{-4} A/cm.² (mm. Hg)²]. A small decrease was observed on the anode fall of potential with increasing pressure (1–5 mm. Hg). The diminution of about 2 volts may be due to the fact that the theory cannot be used for the higher pressures, or to the use of a small cathode and a large anode on which the current density is not equally distributed.

Measurements of the thickness of the anode zone are still uncertain. In H₂ probe measurements give 5 mm., while only 1.5 mm. are found visually. This is because the position of the boundary between the anode glow and the positive column is mainly determined by spectral lines causing the strongest physiological impression, and at the same time possessing the steepest increase of the light intensity in axial direction.

Earlier measurements of the anode fall in various molecular gases were made when the influence of impurity of the gases and of the disturbing effect of probes on the discharge has not been taken into account. Where traces of impurities do not play a very important role, the probe gives an anode fall which is too high by an amount of the order of the volt equivalent of the electron temperature in the positive column.¹ In this way old data (see ref. ⁽³⁾) show that the anode fall is practically independent of the anode material, and at a pressure of the order 1 mm. is nearly equal to the ionization potential, except for oxygen, where negative ions, chemical reactions, etc., may lead to an increase. In oxygen as well as in halogens the thin surface layer attached to the anode, which is built up by the reacting gas, might often simulate a higher anode fall.

In rare gases the anode fall appears to be smaller if a heated cathode is used instead of a cold cathode. Here the anode is immersed in a plasma, the properties of which are mainly determined by the cathode zone. In this case, as well as when the anode is situated near or in the Faraday space of a cold glow discharge (which represents probably a plasma too), the anode mechanism appears to be different from what we had in view. Diffusion processes and the action of excited atoms have been suggested as playing an important part (see refs. ^{(7), (9), (10)}), though here a theory of the anode zone, controllable by experiment, has not yet been given. It has been suggested that in the latter case an anode fall of potential is developed, when the anode current considerably exceeds the random electron current of the plasma.

Further information about the processes involved in the discharge near the anode might be gained by experimental investigations of the distribution of the electrical field strength and of the line spectrum emitted in this region. It is to be expected that, as the velocity of the electrons increases towards the anode, the maximum intensity of the lines will be nearer and nearer to the anode the higher the potential at which the excitation probability of the line concerned is a maximum. This would hold

on the assumption that a considerable number of fast electrons exist which produce excitation near the maximum of the probability curve. By altering only the radius of the discharge tube, the influence on the anode fall of the initial velocity of the electrons which enter the fall zone might be studied and compared with the law of dependance outlined by theory.

The author has much pleasure in expressing his gratitude to Professor Sir J. S. Townsend for the interest he has shown in this investigation, and to Dr. R. W. Lunt for many discussions and help in composing this paper.

References.

- (1) J. S. Townsend, 'Electricity in Gases,' p. 399 (1915).
- (2) A. v. Engel and M. Steenbeck, 'Elektrische Gasentladungen,' vols. 1 & 2 (1932 and 1934).
- (3) J. J. Thomson and G. P. Thomson, 'Conduction of Electricity through Gases,' 3rd ed., vols. 1 & 2 (1928 and 1933).
- (3 a) R. Seeliger, 'Physik d. Gasentladungen,' 2nd ed. (1934).
- (3 b) K. K. Darrow, 'Electrical Phenomena in Gases,' 1932.
- (4) M. Loeb, 'Electrical Discharges in Gases' (1940).
- (4 a) R. Bär, *Handb. d. Phys.* xiv. p. 248 (1927).
- (5) A. Güntherschulze, W. Bär, and H. Betz, *Z. f. Phys.* cix. p. 293 (1938).
- (5 a) A. Güntherschulze, *Z. f. Phys.* xxx. p. 175 (1924).
- (6) S. M. Rubens and J. E. Henderson, *Phys. Rev.* lviii. p. 446 (1940).
- (7) M. J. Druyvesteyn, 'Physica,' iv. p. 669 (1937).
- (8) M. J. Druyvesteyn and M. Penning, *Rev. Mod. Phys.* xii. p. 128 (1940) and xiii. p. 72 (1941).
- (9) K. G. Emeleus, W. L. Brown, H. McN. Cowan, *Phil. Mag.* xvii. p. 146 (1934).
- K. G. Emeleus and O. S. Duffendack, *Phys. Rev.* xlvii. p. 460 (1935).
- (10) R. Seeliger, *Z. f. Phys.* cxv. p. 115 ff. (1940).
- (11) A. Bramley, *Phys. Rev.* xxvi. p. 794 (1925).
- (12) R. W. Lunt, *Roy. Coll. Sci. Journ.* x. p. 123 (1940).
- (13) A. Partzsch, *Z. f. Phys.* xv. p. 287 (1923).

Electrical Laboratory,
Oxford.

XXXVIII. *Some Properties of the Hodograph and their Application in the Construction of the Hodograph in the Problem of External Ballistics.*

By T. LEWIS, Aberystwyth*.

[Received June 18, 1941.]

THE equations of motion of a projectile in a resisting medium can be integrated analytically only for special forms of the law of resistance and, except for the case in which the resistance is proportional to the velocity, the analysis is very complicated and great labour is involved before numerical results are obtained. The equations for an arbitrary law of resistance can be solved numerically with any degree of precision required. But all the calculations have to be repeated for every set of initial conditions and for each value of the ballistic coefficient. The labour of integration would be greatly reduced if the hodograph were known beforehand, or if a quick method of constructing the hodograph were known. Knowledge of the hodograph is equivalent to the knowledge of a first integral of the equations of motion.

In the following paragraphs we prove some properties of the hodograph which appear to have been overlooked by ballisticians. One of these properties leads to a quick and fairly accurate graphical method of constructing the hodograph for an arbitrary law of resistance, provided the ballistic coefficient remain constant.

The properties to be proved are :—

(1) Hodographs are orthogonal trajectories of a certain one-parameter family of curves. These curves are closed curves and the hodographs converge to the limiting point of these curves.

(2) The tangents to all hodographs at points corresponding to a prescribed velocity v pass through a fixed point provided the ballistic coefficient is constant.

(3) The area swept out by the radius vector from the origin to the hodograph is proportional to the horizontal distance travelled by the projectile.

If the one-parameter family of curves mentioned above were constructed for small differences of the parameter we could see at a glance the general behaviour of all hodographs. But in order to construct one hodograph accurately great care would be necessary, and great accuracy

* Communicated by the Author.

could not be guaranteed. Moreover, a different set of curves would have to be constructed for each value of the ballistic coefficient, while the property does not hold for a variable ballistic coefficient.

The second property, however, leads to a quick method of constructing the hodograph. It does not involve the construction of curves other than concentric circles, so that polar graph paper will serve the purpose whatever the value of the ballistic coefficient, provided it is constant. Also, the method can be modified so as to take into account the variability of the ballistic coefficient due to variable atmospheric density.

The third property can be used as a check when the calculations have been completed. Or, given the initial conditions and the velocity at any point of the trajectory, the horizontal displacement can be found directly from the hodograph by integrating the area swept out by the radius vector.

It is obvious that graphical methods cannot rival numerical methods in accuracy. In this connection a quotation from a paper by W. E. Dalby* is relevant: "... tables cannot ensure a greater accuracy than the data from which they are derived, and it is probable that in many cases solutions of problems obtained directly from the ballistic diagram will have an order of accuracy as great as the primitive data, and certainly greater than the order of accuracy of the ballistic constant, since this is influenced by accidental variations of many kinds quite apart from the variation of muzzle velocity produced by the variations in the quality of the propellant and the gravimetric density of the charge."

But whatever the value of this method of constructing the hodograph from a practical point of view, the properties on which it is based are of considerable theoretical interest and they give an insight into the behaviour of the totality of hodographs for a given value of the ballistic coefficient.

Hodographs as Orthogonal Trajectories.

The equations of motion of a projectile may be written in the form

$$\frac{dx'}{dt} = -\frac{x'}{v} cf(v), \quad \frac{dy'}{dt} = -\frac{y'}{v} cf(v) - g, \quad \dots \quad (1)$$

where

$$x' = \frac{dx}{dt}, \quad y' = \frac{dy}{dt}, \quad v^2 = x'^2 + y'^2,$$

and $cf(v)$ is the resistance per unit mass due to the medium when the velocity is v . c , the ballistic coefficient, depends upon several factors, in particular, upon the density of the medium.

* W. E. Dalby, F.R.S., Proc. Roy. Soc. A, vol. xcii. pp. 329-337 (1915-16).

The first order differential equations (1) determine the hodograph. If we now write

$$F(v) = \int^v f(v) dv$$

$$\text{and} \quad \phi(x', y') = cF(\sqrt{x'^2 + y'^2}) + gy', \quad (2)$$

it is clear that equations (1) may be written

$$\frac{dx'}{dt} = -\frac{\partial \phi}{\partial x'}, \quad \frac{dy'}{dt} = -\frac{\partial \phi}{\partial y'}, \quad (3)$$

showing that hodographs are orthogonal trajectories of the one-parameter family of curves

$$\phi(x', y') = \text{constant}, \quad (4)$$

provided c , the ballistic coefficient, remains constant along the trajectory, i.e., provided the density of the medium is uniform.

Let us consider the surface

$$z' = \phi(x', y').$$

The least value of z' is given by

$$\frac{\partial z'}{\partial x'} \equiv -\frac{x'}{v} cf(v) = 0, \quad \frac{\partial z'}{\partial y'} \equiv -\frac{y'}{v} cf(v) - g = 0.$$

This occurs when

$$x' = 0, \quad cf(v) = g, \quad \text{with} \quad y' = -v.$$

If V denotes the limiting velocity in the medium, this minimum value occurs at the point

$$x' = 0, \quad y' = -V, \quad z' = cF(V) - gV.$$

It is easily seen that the surface (4) will form a sort of crater with this point as vertex. All hodographs will have the point $x' = 0, y' = -V$, as limiting point.

As an illustration, let us consider the case

$$f(v) = kv, \quad F(v) = \frac{1}{2}kv^2, \\ \phi(x', y') = \frac{1}{2}ck\{x'^2 + (y' + V)^2 - V^2\}, \quad V = g/ck.$$

The curves $\phi = \text{constant}$ are concentric circles with the point $(0, -V)$ as centre, and the hodographs are radii of these circles.

In general the curves $\phi = \text{constant}$ are closed curves, symmetrical with respect to the y' -axis, having the point $(0, -V)$ as limiting point. Hence, if the curves are drawn with sufficient density and on a fairly large scale, the hodographs can be constructed with fair accuracy.

Although of considerable theoretical interest this method has serious disadvantages from a practical point of view. In the first place, it is no

use at all for high elevation trajectories because it takes no account of the change in atmospheric density. Secondly, a set of curves would have to be drawn for each value of the ballistic constant even for low trajectories, so that the labour involved would be very considerable, if not prohibitive.

A Tangent Property of Hodographs.

A normal to the curve $\phi = \text{constant}$ is the tangent to a hodograph passing through the same point. Its equation is

$$(\eta - y') \frac{\partial \phi}{\partial x'} = (\xi - x') \frac{\partial \phi}{\partial y'}, \quad \dots \dots \dots (5)$$

where (ξ, η) are running co-ordinates.

This tangent cuts the y' -axis at the point

$$\begin{aligned} \eta &= \left(y' \frac{\partial \phi}{\partial x'} - x' \frac{\partial \phi}{\partial y'} \right) / \frac{\partial \phi}{\partial x'}, \\ &= \left\{ y' \left(\frac{x'}{v} cf(v) \right) - x' \left(\frac{y'}{v} cf(v) + g \right) \right\} / \frac{x'}{v} cf(v), \\ \text{i. e.,} \quad \eta &= -gv/cf(v). \quad \dots \dots \dots (6) \end{aligned}$$

This result shows that the tangents to all hodographs at the points corresponding to the velocity v pass through a fixed point on the y' -axis provided c , the ballistic coefficient, remains constant. This property leads to a quick method of constructing the hodograph, which is free from the limitations of the first method proposed. It can be extended so as to allow for the variation of density and it does not require the construction of a different set of curves for each value of c ; all that is needed is polar graph paper and tables of values of η , which can be calculated from the tables for $f(v)$. The method will be described later.

The Area swept out by the Radius Vector to the Hodograph and its relations to the Horizontal Range of the Projectile.

Multiplying the first of equations (1) by $-y'$, the second by x' and adding we get the equation

$$x' \frac{dy'}{dt} - y' \frac{dx'}{dt} = -gx',$$

which is equivalent to

$$x' dy' - y' dx' = -g dx. \quad \dots \dots \dots (7)$$

On integration we get

$$\int (x' dy' - y' dx') = -gx. \quad \dots \dots \dots (8)$$

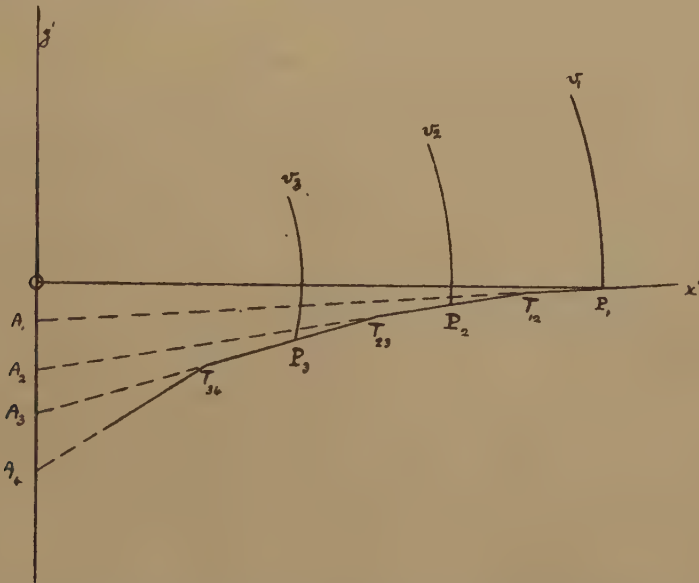
The left-hand side is twice the area swept out by the radius vector from the origin to the hodograph. It follows that when the hodograph has been constructed the horizontal displacement of the projectile corresponding to any point of the hodograph can be found from it directly without further integration. In particular, it can be seen that the total horizontal distance is finite, because all hodographs converge to the point $(0, -V)$. Moreover, the property is valid for a variable ballistic coefficient.

Construction of the Hodograph when the Ballistic Coefficient is Constant.

In order to construct the hodograph we have first to calculate the function

$$\psi(v) = v/f(v) \quad \dots \dots \dots (9)$$

Fig. 1.



for the different values of v , $f(v)$ being known from the resistance tables. We have already seen that when the velocity is v , the tangent to the hodograph passes through the point $(0, \eta)$, where

$$\eta = -(g/c)(v/f(v)) = -(g/c)\psi(v). \quad \dots \dots \dots (11)$$

The diagram (fig. 1) is illustrative only, and for this reason the intervals $\Delta v = v_1 - v_2 = \text{etc.}$ are comparatively large.

We next select two perpendicular radii on polar graph paper as axes of x' and y' , and mark off points $A_1, A_2 \dots$ along the negative y' -axis

at distances $g/c(\psi_1, \psi_2 \dots)$ from the origin, corresponding to values $v_1, v_2, \dots$ of v , as shown in the figure. If P_1 is the initial point of the hodograph, corresponding to $v=v_1$, a ruler is laid along $P_1 A_1$, which is tangent to the hodograph at P_1 , and a short segment $P_1 T_{12}$ is drawn, where T_{12} is the point where the tangent intersects the circle $v=v_1-\frac{1}{2}\Delta v$, Δv being the difference v_1-v_2 . Next, the ruler is laid along $T_{12} A_2$, which is (approximately) tangent to the hodograph at the point where it meets the circle $v=v_2$. It will be shown later that the actual distance between this line and the curve is a small quantity of third order, Δv being regarded as a small quantity of first order. The segment $T_{12} T_{23}$ is drawn along this line, where T_{23} is on the circle $v=v_2-\frac{1}{2}\Delta v$. This segment cuts the circle $v=v_2$ at a point P_2 , which is (approximately) on the hodograph concerned. Next, a segment $T_{23} T_{34}$ is drawn along the join of $T_{23} A_3$. The intersection of this section with the circle $v=v_3$ gives the point P_3 , which is (approximately) on the hodograph. This process is repeated until the diagram is completed.

In order to get satisfactory results the diagram must be drawn to a fairly large scale. Suppose, for example, 1000 m/s. is represented by 1 metre, and the circles (it is sufficient to have semi-circles) are drawn at intervals of 3 mm., so that each increment represents 3 m/s., which is roughly the interval for which the resistance function $f(v)$ is tabulated. The points $T_{12}, T_{23} \dots$ can be found very quickly as follows:—The polar diagram is fixed to a drawing-board which is set in a vertical plane. A rigid scale is fixed along the y' -axis and carries a slide with a fine needle projecting normally to the board. One end of a fine thread is fixed at the initial point P_1 of the hodograph, and the thread is passed over the projecting needle on the slide and a small weight is attached to the other end so as to keep it taut. The slide is now moved until the needle is directly over the point A_1 , so that the thread is tangent to the hodograph at P_1 . A fine needle is now fixed at the point T_{12} on the circle $v=v_1-\frac{1}{2}\Delta v$, on the concave side of the hodograph. The slide is now moved so that the projecting needle is over A_2 , so that the thread is (approximately) tangent to the hodograph, where it meets the circle $v=v_2$, and a second needle is fixed at the point T_{23} , etc. When completed the thread will form a polygon whose sides are (approximately) tangent to the hodograph at the points $P_1, P_2 \dots$, where they intersect the circles $v=v_1, v_2 \dots$

Determination of the Trajectory from the Hodograph.

From the first of equations (1) we get

$$cdt = -\frac{vdx'}{x'f(v)} = -\frac{\psi(v)dx'}{x'} \quad \dots \dots \dots (12)$$

When the hodograph has been constructed v , and hence $\psi(v)$, will be

known for every value of x' . Hence, by integration, numerical or graphical, we can obtain t as a function of x' , i. e.,

$$ct = - \int_{x_1}^{x'} \frac{\psi(v) dx'}{x'} \dots \dots \dots (13)$$

This operation also gives us x' for every value of t .

From the equation

$$\frac{dx}{dt} = x'(t)$$

we get, by integration, numerical or graphical,

$$x = \int_0^t x'(t) dt, \dots \dots \dots (14)$$

i. e., x as a function of t . This value could be checked by comparing it with the area swept out by the radius vector in the hodograph diagram.

y' and y can be found as functions of t in precisely the same way, or, since the hodograph gives y' as a function of x' , which is now known as a function of t , y' is known as a function of t , and

$$y = \int_0^t y'(t) dt. \dots \dots \dots (15)$$

(14) and (15) determine the trajectory completely.

Variation of Atmospheric Density taken into account.

It has been shown that the tangent to a hodograph at a point corresponding to the velocity v always passes through the point $(0, \eta)$, where

$$\eta = -(g/c)\psi(v).$$

If we now assume that the density of the atmosphere varies with the height, y , above sea-level, c will be a function of y , i. e.,

$$c = c_0 \chi(y), \text{ where } \chi(0) = 1,$$

so that

$$\eta = -(g/c_0) \frac{\psi(v)}{\chi(y)}. \dots \dots \dots (16)$$

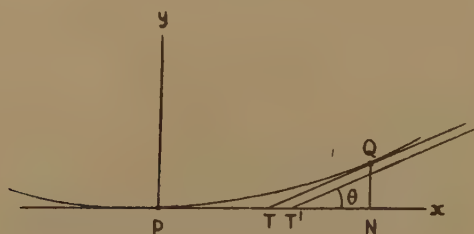
It is now impossible to complete the hodograph from the initial data independently of the trajectory. Nevertheless, the problem can be solved completely with roughly the same labour as for a constant c . The only difference is that y' and y have to be calculated step by step as the hodograph is constructed, and the appropriate value of y substituted on the left-hand side of (16) in order to find the next value of η . If desired, x' and x can be calculated after completing the hodograph, or they can be calculated simultaneously. We thus see that there is no new principle

involved when the density varies in this way. In fact the method can be applied when the density is a function of the two variables x and y . However, in the case of variable density the method loses its elegance, and, owing to the necessity for carrying on all the integrations at the same time, loss of speed of operation will be inevitable.

An Estimate of the Accuracy of the Construction.

In the solution of a problem involving graphical construction errors arise from two distinct sources. In the first place, error arises because there is a limit to the sensitivity of the instruments used, and to this must be added the personal error. Only an experienced draughtsman can estimate the errors from these sources, but their effect can be reduced by increasing the scale of the diagram (fig. 1). In the second place, there is the error due to working with finite increments of the independent variable instead of infinitesimal increments. A rough estimate of the error arising from this cause may be calculated as follows:—

Fig. 2.



Let the tangent to a curve at P be taken as x -axis and the normal as y -axis. In the neighbourhood of this point the equation of the curve is of the form

$$y = ax^2 + O(x^3).$$

The equation of the tangent at a neighbouring point is

$$\eta - ax^2 = (\xi - x)2ax + O(x^3).$$

This tangent cuts the x -axis at the point $T(\xi, 0)$, where

$$\xi = \frac{1}{2}x + O(x^2).$$

Hence, neglecting small quantities of second order,

$$PT = TN = TQ,$$

where N is the projection of Q on the tangent at P . Suppose T' is a point near T such that $TT' = O(x^2)$, and a line is drawn through T' parallel

to the tangent TQ, the perpendicular distance between these parallel lines is $TT' \sin \theta$, where θ , the slope of the tangent, is a small quantity ($=2ax + O(x^2)$), i.e., this distance is a small quantity of third order.

In the construction of the hodograph, described in a previous paragraph, $P_1T_{12}=T_{12}P_2$ to the first order, so the error in the position of the point P_2 is due to an error of second order in the position of T_{12} , the point of intersection of neighbouring tangents. It then follows from the above investigation that the error in the position of P_2 is of the third order. An additional error of the same order of magnitude arises at each stage of the construction. Since these errors are cumulative, the resultant error in the position of a point P_n , which is at a finite distance from P_1 , will be of the second order—a highly satisfactory result.

When the contact of the curve with its tangent is of a higher order, as at points of inflection, for example, the above argument does not hold, but owing to this closer contact the error in our construction at such points will still be of third or higher order, and, in any case, there will be only a finite number of such points on a single hodograph, so that they cannot upset the construction as a whole. Points of inflection are of interest because they correspond to a minimum value of the speed in the trajectory. At such a point the hodograph is tangent to the circle $v=v_{\min}$. In order to locate a point of inflection accurately it is advisable to make the intervals Δv smaller, because the corresponding increment of arc of the hodograph is of order $\sqrt{\Delta v}$.

Owing to the uncertainty in the value of the ballistic constant, it is of greater importance to be able to calculate accurately the *difference* in the elements of two trajectories due to a change in the initial conditions. If the difference in the initial conditions is a small quantity of order Δv , it follows that the error in the difference at any stage, due to the *method* of constructing the hodograph alone, is of the third order. So considered as a graphical method it appears to be highly satisfactory.

XXXIX. *The World Mapped in Complementary Segments.*

By A. F. DUFTON, M.A., D.I.C.*

[Received September 1, 1941.]

*History and SCIENCE our playthings are : what an untold
 Wealth of inexhaustive treasure is stored up for amusement !
 Shall the amass'd Earth-structure appeal to me less than in early
 'Childhood an old fives-ball, whose wraps I wondering unwound . . . ?*

R. BRIDGES

OF the sphere $x^2 + y^2 + z^2 = r^2$, it may be remarked that if the half for which the abscissa x is positive is cut by the two parallel planes $2z^2 = r^2$, and the half for which the abscissa is negative by the planes $2y^2 = r^2$, the surface divides into two similar, equal, segments.

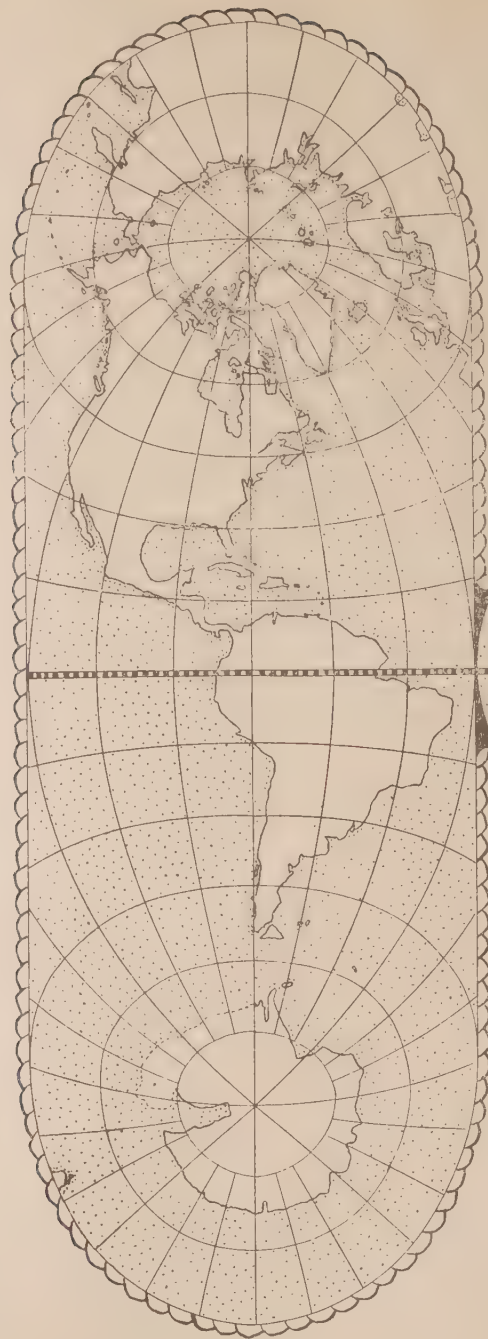
These segments provide a suitable partition of the surface of the Earth when it is desired to exhibit the World in two complementary maps and, for this purpose, appear to some advantage when compared with the usual division along a great circle. The mapping of a hemisphere, involving as it does the representation of an area of $2\pi r^2$ within a perimeter of $2\pi r$, is not accomplished without considerable distortion.

In the annexed maps, the World is shown by an *equal area* projection. For the non-polar segment, the parallels of latitude are made rectilinear and the scale of longitude aptly proportioned to the cosine-of-the-latitude augmented by two : the graticule of the polar map is derived from that of the equatorial one.

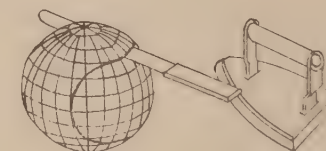
It is clear that a variety of complementary segments can be cut, and for certain purposes an oblique projection may be preferred to a presentation symmetrical with respect to the equator.

The King's Lodge,
 Hunton Bridge, Hertfordshire.

* Communicated by the Author.



The World
newly presented
in two complementary maps



XL. *The Steady Flow of Heat from certain Objects buried under Flat Air-cooled Surfaces.*—Phil. Mag. p. 471 (June 1941).

To the Editors of the Philosophical Magazine.

GENTLEMEN,—

DR. BICKLEY has drawn my attention to the fact that, since

$$\int_0^{\infty} \{\cos a\theta / \cosh \theta\} d\theta = (\pi/2) \operatorname{sech} (\pi a/2),$$

my expression $f(\lambda, \pi/2l)$, which is of series form, can be replaced by $(\pi/4) \operatorname{sech} (\lambda l)$ in equation (44), with a resulting simplification in the evaluation of the integral.

Yours faithfully,
F. H. SCHOFIELD.

August 12, 1941.

XLI. *Miscellaneous.**Project for the Computation of Mathematical Tables.*

The Mathematical Tables Project is now nearly four years old. It will be recalled that its object is to compute and publish tables of mathematical functions, in whatever science they may be required. Thus the first volume (which was published in 1939) contained the first ten powers in full of all integers up to 1000, a table in process of reproduction gives section moduli and moments of inertia of use to naval architects, whilst yet others include tables of use to geographers, statisticians, or astronomers.

It is natural to compare the scheme with that of our own British Association Tables Committee, which also includes tables of use in pure mathematics, statistics, and astronomy. Tables, even extensive ones, have been calculated before, and are witnesses to the industry of such men as Gauss, Barlow, and Glaisher; but it is only in recent years that the development of the use of calculating machines, and the new methods of computation based on them, have made it possible to carry out calculations fairly quickly, and, more important, to assign a high credibility to the statement "It is believed that no errors remain." This new state of affairs has, of course, been responsible for the close approach in time between the activities of the two bodies. It might at first sight seem that there would be wasteful overlapping, but on examination this is not found to be so. To take again the first volume, with x^n for n up to 10 and x up to 1000; this was followed the next year by the British Association's volume 9, giving x^n up to $n=12$ and $x=1099$, with extensions both to higher x and lower n , and to higher n and lower x . Its preface was able to record that no errors were found in the American table. The Mathematical Tables Project has now completed, but not published, a table of x^n where n is less than unity, and has under consideration a table of x^{-n} for $n=1$ to 10 and $x=1$ to 1000.

Other fundamental tables planned or published include e^x to 15 decimals, sin and cosine to many figures, the exponential and probability integrals, natural logarithms of integers up to 100,000 and arctan x up to 10,000 (a very useful table). Then there are various Bessel functions, some for complex arguments, the integrals of $x^k \sin n\pi x$ and $x^k \cos n\pi x$ (which we believe have never been tabulated before), and the associated Legendre functions. Finally, there is a possibility of extensive tables of elliptic functions. Altogether there are 43 volumes at various stages, of which eight have actually appeared and another eight will shortly do so.

The project is sponsored by the Bureau of Standards, Washington.

XLII. Notices respecting New Books.

The Identification of Molecular Spectra. By Drs. R. W. B. PEARSE and A. G. GAYDON. [Pp. viii.+221.] (London: Chapman & Hall, Ltd., 1941. Price 42s. net.)

In every previous book on molecular spectra the main subject of the text is the analysis of band systems and individual bands of various types in accordance with the quantum theory, with a view to the derivation of numerical constants and other information concerning electronic states and vibrational and rotational levels of the molecules, and the observational data presented in accompanying tables, diagrams and spectrograms are arranged in a manner best suited to this theme. To all such books the present work, which is quite different in both purpose and plan, may be regarded as a very valuable supplement.

As its title implies, its object is to guide the practical spectroscopist in the often troublesome and tedious task of identifying the bands in a composite spectrogram and of avoiding the old troubles and pitfalls. The need for a guide of this kind is only too evident from the frequency with which wrong assignments have found their way into print, as a result of the tendency to accept as a desired new band system a part of an old system which is not immediately recognized by the investigator, but which turns out to be, perhaps, the CO Third Positive, the CO Fourth Positive or the CS ultra-violet system, arising from impurity in a discharge tube, or the Schumann-Runge system of O₂ produced by absorption by air in the optical path.

The book is admirably designed to fulfil its special purpose. It consists mainly of two sections: (i.) a table of data for 1532 persistent band-heads arranged in order of wave-length from 10603 Å. to 2006 Å.—corresponding to, say, Kayser's 'Tabelle der Hauptlinien der Linienspektren aller Elemente'; and (ii.), the much larger section, a set of lists of data for separate band systems arranged in alphabetical order of the neutral and ionized molecules (254 diatomic and about 34 triatomic and polyatomic) responsible for them. From a comparison of the wave-length and degradation of an unrecognized strong band with the data listed in (i.) we obtain a clue to the identity of the system and its emitter; then, by reference to the particulars for that system in (ii.), we obtain a check from the presence or absence of other bands of the system. This process of double reference is now repeated for bands of other systems of the same molecule and of any other molecule having an element in common with it; *e. g.*, having identified a system of C₂ and one of N₂, we look for other systems of C₂ and N₂ and for systems of CN, CO and NO (since air or O₂ may not be entirely absent from the source).

The importance of atomic lines in this work is emphasized, and a table of about 400 persistent lines of elements in alphabetical order of their chemical symbols is provided. In view of the utility and rapidity of direct comparison as a means of identification, six plates of well chosen and beautifully reproduced band spectrograms are included. In addition, there are two plates of comparison spectra (Fe, Cu, Hg, Ne) for use in wave-length measurement. The

work concludes with a ten-page section containing valuable practical hints on matters which are often troublesome to the beginner and seldom treated in textbooks, and a six-page index of authors and subjects.

We may be sure that copies of this excellent book will soon be acquired by spectroscopic laboratories throughout the civilized world as an indispensable item of their practical equipment.

W. J.

Spectrochemical Abstracts, Vol. II, 1938-9. By E. H. S. VAN SOMEREN.
[Pp. 38.] (London: Adam Hilger, Ltd., 1941. Price 5s., post free.)

MR. F. TWYMAN'S Volume I of these Abstracts covered the years 1933-7; its sequel covers the next two years, but includes also a few important earlier papers and some books published in 1940. The first half of the booklet is occupied by a bibliography, arranged in alphabetical order of the authors' names, and the second half contains 165 short abstracts of papers and some brief notes on books, reviews, charts and tables. The numbering of the abstracts is consecutive with that in Volume I. Of the papers abstracted, 48 are concerned with metals and alloys, 16 with biological materials, 16 with non-metallic solids, 10 with liquids and gases, and 7 with glass and pottery. Apparatus used in such investigations is described in about 50 of the papers abstracted, technique in about 40 and theory and principles in about 20. This brief analysis of the contents indicates both the increasing extent to which spectrographic methods are being employed to the analysis of materials and also the usefulness of this booklet to all who are engaged upon investigations of this kind.

W. J.

*[The Editors do not hold themselves responsible for the views
expressed by their correspondents.]*

XLIII. Viscosity-Temperature Function of Liquids.

By A. H. NISSAN, Ph.D., B.Sc., A.M.Inst.Pet.*

[Received March 31, 1941.]

It has been shown ^{(1), (2)} that when the viscosity, or its logarithm, of the normal paraffins is plotted against the variable $\left(\frac{T_B}{T}\right)$, a single curve is obtained for all members, where

T = absolute temperature of measurement,

T_B = absolute boiling temperature under atmospheric pressure.

Fig. 1 shows this curve, which covers the whole temperature range under which the n -paraffins from C_4 to C_{18} exist as liquids. Similarly, of course, the plot of the logarithm of viscosity against the reciprocal function $\left(\frac{T}{T_B}\right)$ yields a single curve for the normal paraffins. As this latter curve is the basis of certain other studies, the equation of the curve becomes of interest.

To find the equation,

$$\log \mu = f\left(\frac{T}{T_B}\right),$$

$f\left(\frac{T}{T_B}\right)$ is assumed to be a power series of $\frac{T}{T_B}$, thus

$$\log \mu = K + K_1\left(\frac{T}{T_B}\right) + K_2\left(\frac{T}{T_B}\right)^2 + K_3\left(\frac{T}{T_B}\right)^3 \dots$$

In order to find out how many terms are necessary to satisfy the equation accurately, the equation was solved for the mean curve passing through the points for all the hydrocarbons tested, taking terms up to and including $\left(\frac{T}{T_B}\right)^3$. Thus, by taking the values of $\log \mu$ at values of $\frac{T}{T_B}$ of 0.6, 0.7, 0.8, and 0.9, it is found that

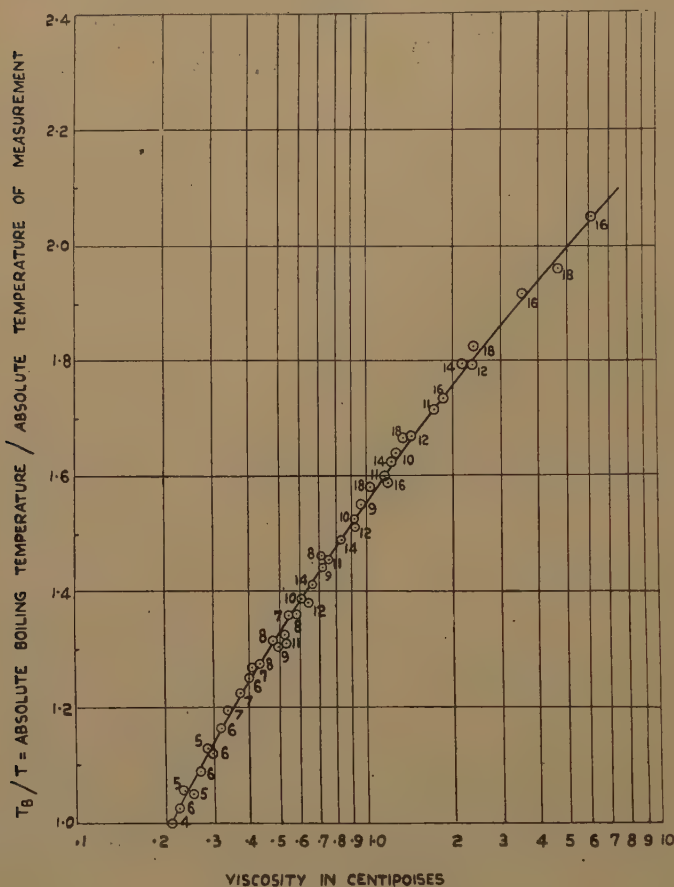
$$\log \mu = 6.283 - 18.572 \left(\frac{T}{T_B}\right) + 17.505 \left(\frac{T}{T_B}\right)^2 + 5.900 \left(\frac{T}{T_B}\right)^3,$$

* Communicated by the Author.

where μ = viscosity in centipoises as read from the mean curve representing the data of all the normal paraffins studied.

Table I. gives a comparison of the calculated and mean experimental values of the viscosity found by this method.

Fig. 1.



In this connexion it is of interest to note that viscosity values at the boiling points of normal paraffins available in the literature are as follows:—for *n*-octane, 0.198; *n*-heptane, 0.199; *n*-hexane, 0.204, and for *n*-pentane, 0.200, giving an average of 0.200 centipoises. The normal paraffin curve extrapolates to 0.21 ± 0.005 centipoises.

To prove that the equation applies to any particular liquid or set of liquids would necessitate similar calculations to those involved in computing Table I. This method is lengthy and a simpler graphical method

was employed. It will be seen later that while the graphical method is as rigorous as the detailed calculations would be, the graphs bring forth certain points which are of importance from physico-chemical view-points because of certain simplifications in utilizing the general equation,

$$\log \mu = K - K_1 \left(\frac{T}{T_B} \right) + K_2 \left(\frac{T}{T_B} \right)^2 - K_3 \left(\frac{T}{T_B} \right)^3 + \dots$$

Fig. 2 illustrates the procedure. The curve for the *n*-paraffins was plotted, using logarithmic scale for the abscissa and natural scale of $\frac{T}{T_B}$ for the ordinates. A straight line was drawn joining points (0.21, 1.000) and (4.4, 0.500) and this line was then assumed to represent the viscosity-

TABLE I.

$\frac{T}{T_B}$	μ Experimental.	μ Calculated.	$\Delta\mu$ Error.
0.500	4.4	4.3	0.1
0.550	2.43	2.45	-0.02
0.600	1.47	1.47	0.00
0.650	0.970	0.980	-0.010
0.700	0.685	0.686	-0.001
0.750	0.513	0.513	0.000
0.800	0.405	0.405	0.000
0.850	0.330	0.332	-0.002
0.900	0.278	0.279	-0.001
0.950	0.239	0.242	-0.003
1.000	0.21	0.207	0.003

N.B.—The values at $\left(\frac{T}{T_B} \right) = 0.500$ and 1.000 were taken only to two significant figures, as there is uncertainty in experiments at or near both the melting and boiling points.

temperature relationship of the *n*-paraffins on a chart, the *x*-axis of which was in logarithmic scale and the *y*-axis of which was a temperature scale, the equation of which was

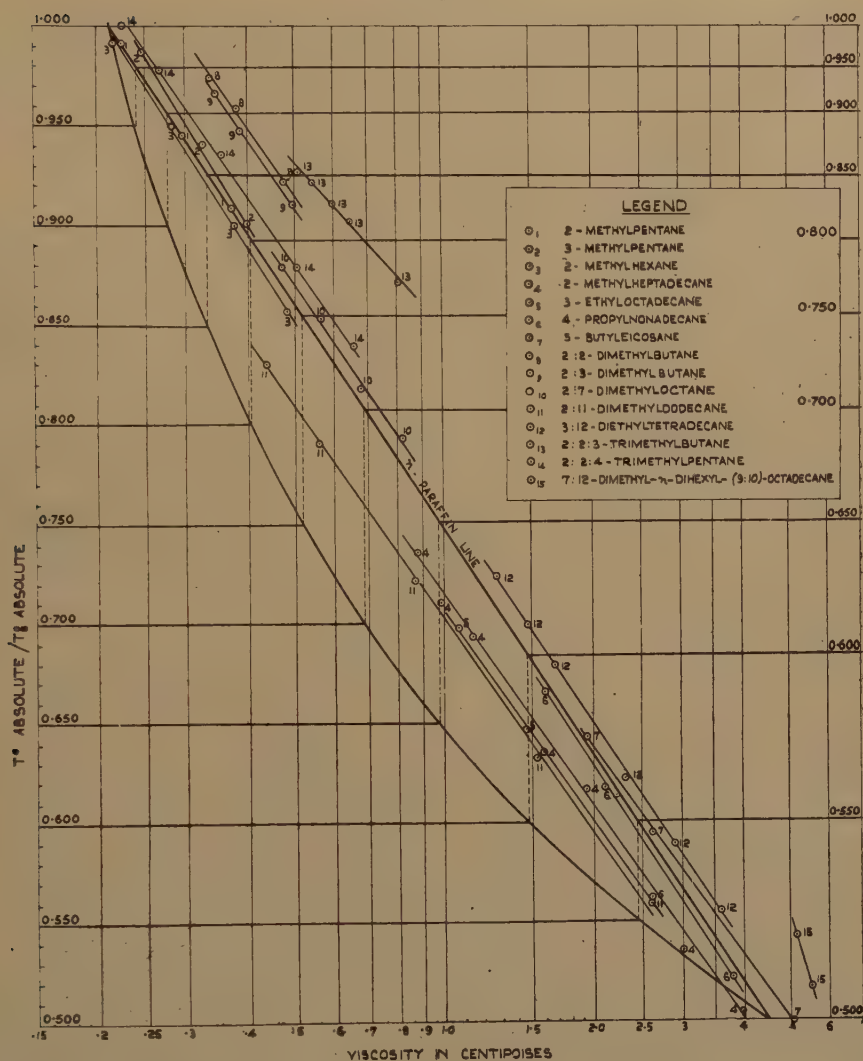
$$y = a - b \left(\frac{T}{T_B} \right) + c \left(\frac{T}{T_B} \right)^2 - d \left(\frac{T}{T_B} \right)^3 + \dots$$

In other words an artificial scale for temperature was used, so that it yields a straight line for the log. viscosity-temperature relation of *n*-paraffins. The method of constructing the scale, which is shown on

$$=a'-b'\left(\frac{T}{T_B}\right)+c'\left(\frac{T}{T_B}\right)^2-d'\left(\frac{T}{T_B}\right)^3+\dots$$

$$\text{But } \log \mu = K - K_1\left(\frac{T}{T_B}\right) + K_2\left(\frac{T}{T_B}\right)^2 - K_3\left(\frac{T}{T_B}\right)^3 + \dots$$

Fig. 3.



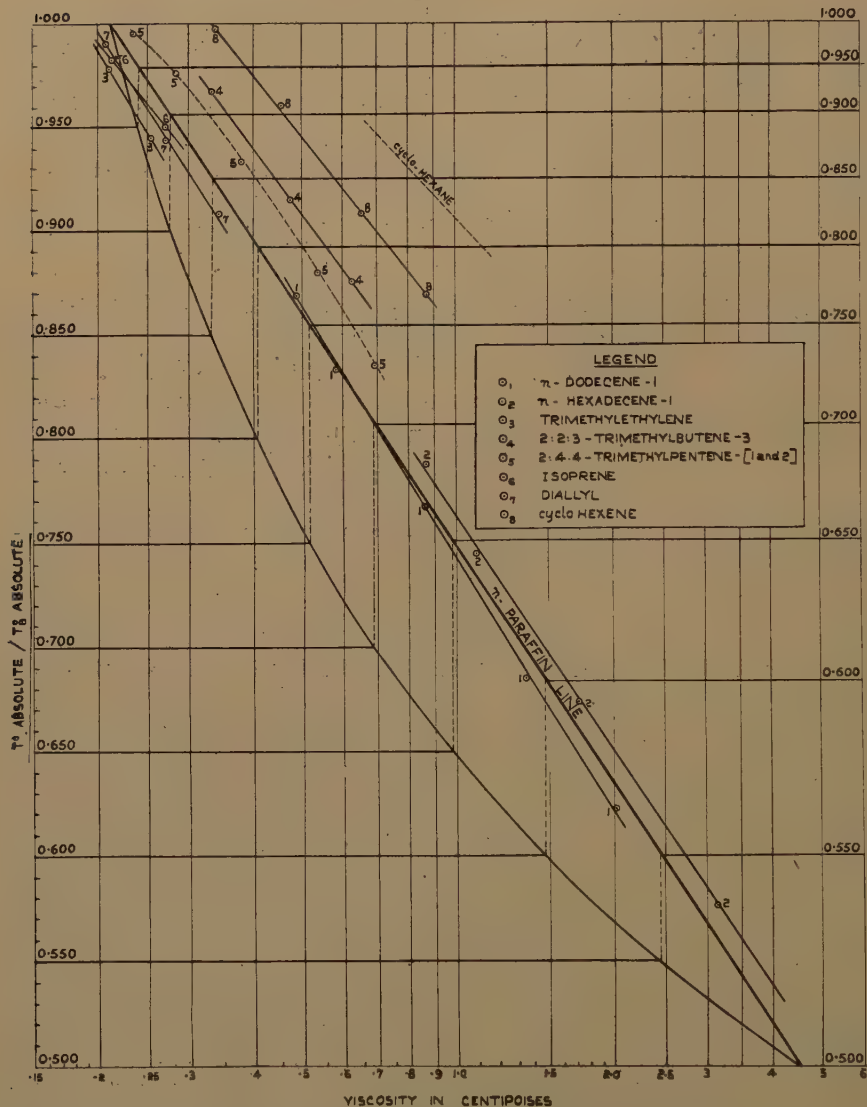
Thus, for the straight line to represent the *n*-paraffin behaviour with regard to viscosity change with temperature,

$$\frac{b'}{K_1} = \frac{c'}{K_2} = \frac{d'}{K_3} = \text{constant.}$$

Similarly, for any straight line drawn on this artificial scale an equation may be written in the form

$$\log \mu = K' - K_1' \left(\frac{T}{T_B} \right) + K_2' \left(\frac{T}{T_B} \right)^2 - K_3' \left(\frac{T}{T_B} \right)^3 + \dots$$

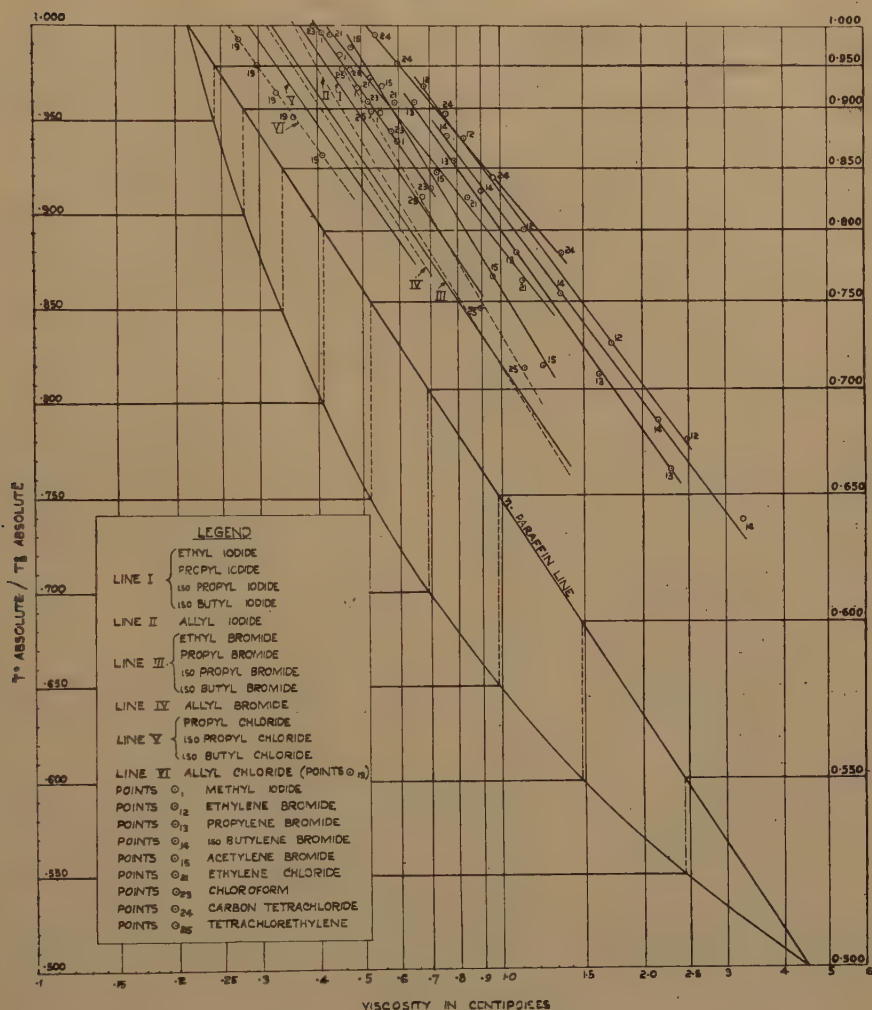
Fig. 4.



Thus any liquid, the viscosity-temperature relationship of which yields a straight line on this chart, may be said to obey this general law.

Charts constructed in this manner have been plotted for the liquids shown on figs. 2, 3, 4, 5, 6, 7, and 8. The vertical axis was 500 mm. in length, and to ensure accuracy of plotting no interpolation was attempted on the artificial scale but each point was plotted by first taking the value

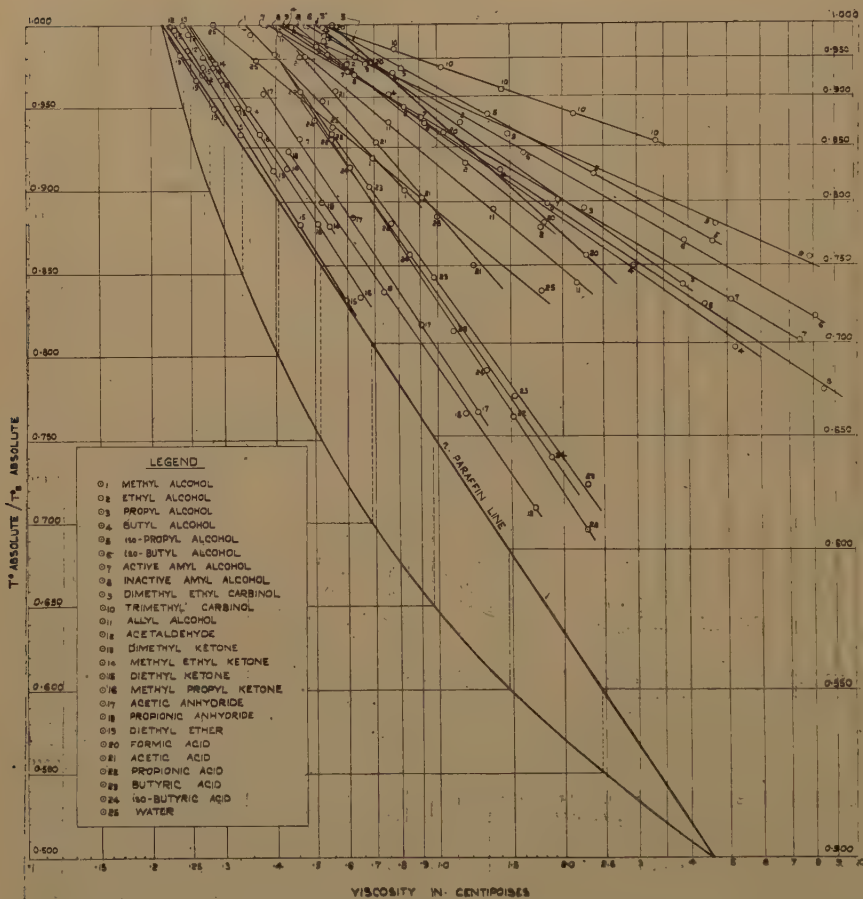
Fig. 5.



of $\left(\frac{T}{T_B}\right)$ on the natural scale on the left, moving to the right until the *n*-paraffin curve is met, erecting an ordinate at this intersection to meet the straight "*n*-paraffin line" and drawing a horizontal line at this

point of intersection to show the value of $\left(\frac{T}{T_B}\right)$ on the artificial scale. The intersection of this horizontal line with the $\log \mu$ value on the x -axis determined the point.

Fig. 6..

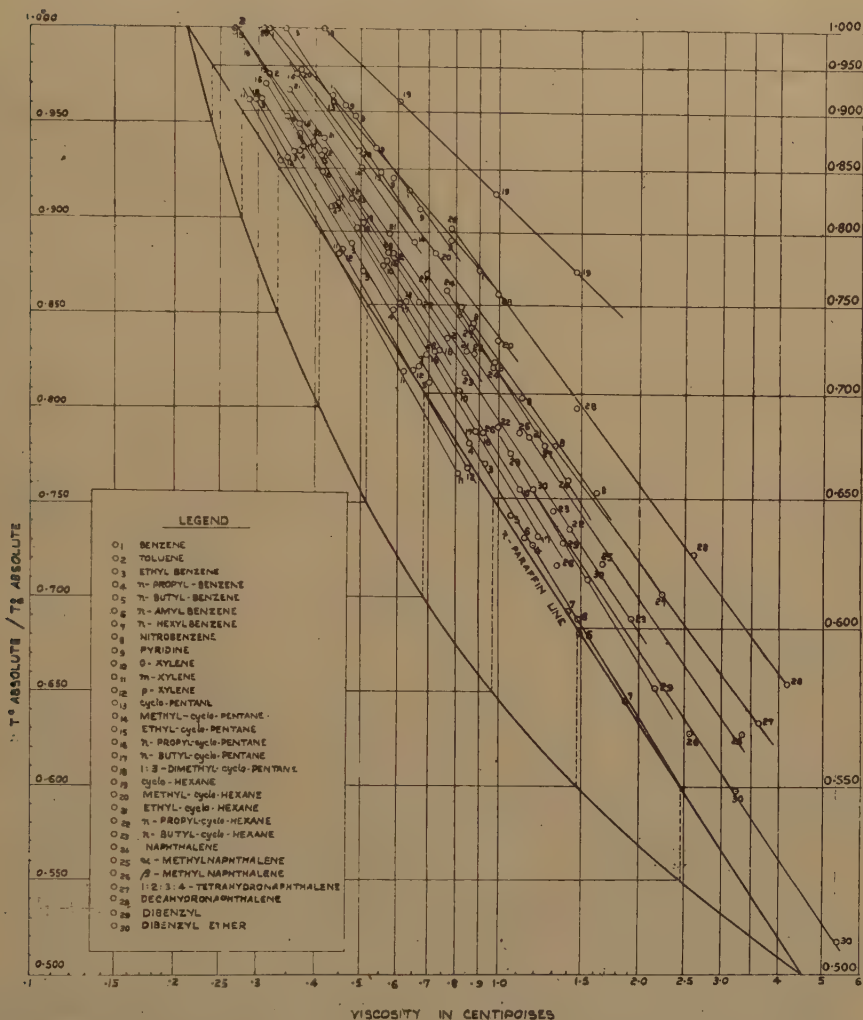


It will be seen that straight lines were obtained for all the liquids which were tested (the majority of which possessed more than 4 experimentally found values) excepting:

- (1) 2 : 4 : 4-Trimethylpentene-(1 and 2), which is a mixture of 2 isomers and thus is not a pure liquid (fig. 4) ;
- (2) Sodium chloride, the boiling point of which was not known accurately and was taken 1686° abs. as an approximate value (fig. 8).

Over a hundred liquids composed of organic and inorganic compounds and elements were tested. These were chosen to yield contrasting types of liquid, *i. e.*, unassociated and inert, against associated to various

Fig. 7.

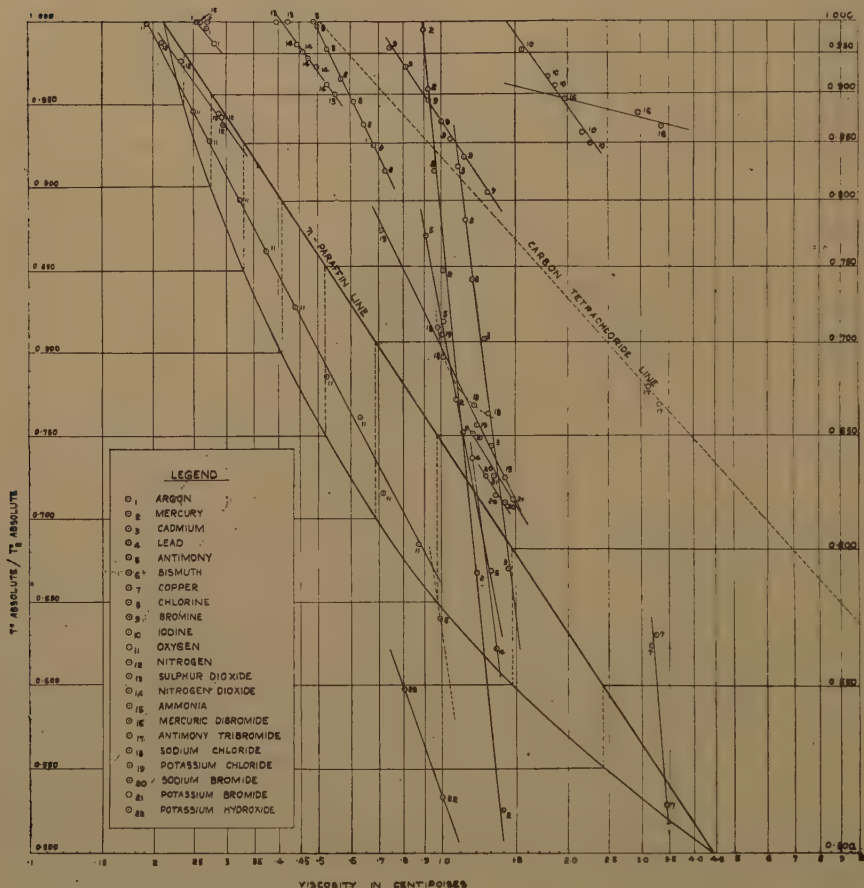


degrees, or dissociated ionic fused salts and metals; liquids of ideal spherical symmetry such as argon, against liquids of various degrees of anisotropy and polarity. Further, the choice was made in order that liquids representing some of the principal organic and inorganic structures

and radicals were tested. All these liquids are then governed by the general equation

$$\log \mu = K' - K_1' \left(\frac{T}{T_B} \right) + K_2' \left(\frac{T}{T_B} \right)^2 - K_3' \left(\frac{T}{T_B} \right)^3 + \dots$$

Fig. 8.



Evidently, unless every liquid is tested, it cannot be stated that an equation based on such an empirical derivation is general to all liquids; but considering the extreme variation in the nature of the liquids tested it is suggested that the equation applies to all pure liquids, provided they do not change identity (or boiling point) with temperature change, *e. g.*, sulphur.

Again, as T_B is a constant for any liquid, then

$$\log \mu = K'' - K_1'' (T) + K_2'' (T)^2 - K_3'' (T)^3 + \dots$$

By using $\left(\frac{T}{T_B}\right)$, however, many of the homologous families yield a single line for all the members which is called the "specific viscosity curve" or the "family viscosity curve." The significance of these curves was already studied in detail in another paper from the point of view of chemical structure and molecular shape.

Reverting to the equation of the paraffins, it was seen that

$$\log \mu = 6.283 - 18.572 \left(\frac{T}{T_B}\right) + 17.505 \left(\frac{T}{T_B}\right)^2 - 5.900 \left(\frac{T}{T_B}\right)^3 +,$$

and that any liquid yielding a straight line on these charts was governed by an equation :

$$\log \mu = K' - K_1' \left(\frac{T}{T_B}\right) + K_2' \left(\frac{T}{T_B}\right)^2 - K_3' \left(\frac{T}{T_B}\right)^3 + \dots$$

in such a manner that

$$\frac{K_1'}{18.572} = \frac{K_2'}{17.505} = \frac{K_3'}{5.900} = \text{constant.}$$

Thus,

$$\log \mu = A - B \left[18.572 \left(\frac{T}{T_B}\right) - 17.505 \left(\frac{T}{T_B}\right)^2 + 5.9 \left(\frac{T}{T_B}\right)^3 \right],$$

where A and B are constants characteristic of the liquid, A being connected with the shape of the molecule and B with its state of association⁽¹⁾. This formula makes it possible to calculate the viscosity-temperature relationship of any pure liquid, the boiling point of which is known, from the values of its viscosity at only two temperatures. A further simplification in the actual calculation is as follows :—

For the paraffin, $A=6.283$ and $B=1$, thus the value of the series in the brackets for any value of $\frac{T}{T_B}$ is given by

$$6.283 - \log [\mu_P]$$

where

$[\mu_P]$ = viscosity value as given by the curve for n -paraffin, at $\frac{T}{T_B}$, the same as that of the liquid under test.

Thus, for any liquid,

$$\log \mu = A - B[6.283 - \log [\mu_P]],$$

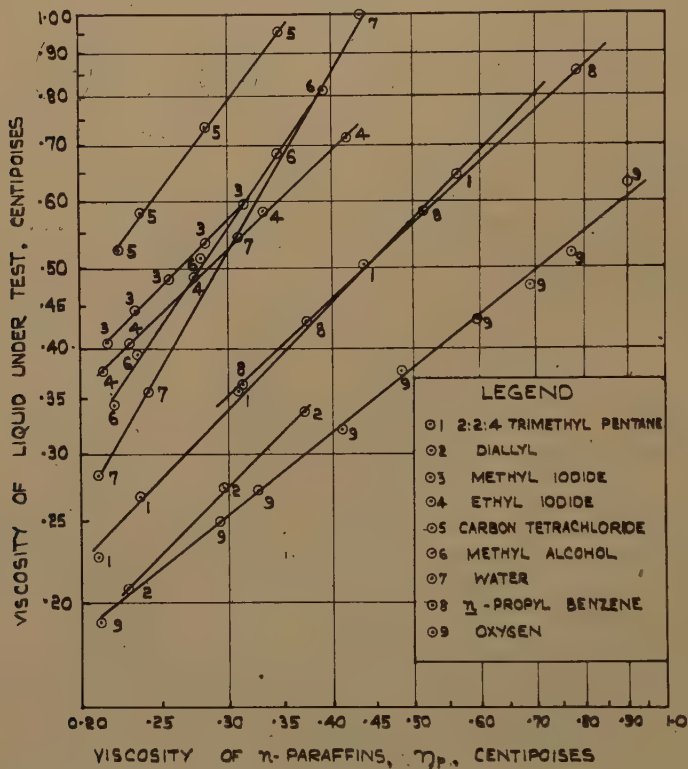
or

$$\log \mu = C + B \log [\mu_P].$$

As the values of $[\mu_p]$ are known to the third significant figure from 0.55 to 0.95 T_B with a maximum error of 1 per cent., then the viscosity of any liquid can be calculated, using the chart, for wide variations in temperature from only two experimental values.

Further to the proof furnished by figs. 2-8 inclusive, this equation is seen to apply in fig. 9, where representative liquids of various types are included. As $\log [\mu_p]$ is thus useful in calculating viscosity-temperature

Fig. 9.



relationships of other liquids, a table is presented giving values of $[\mu_p]$ for different values of $\frac{T}{T_B}$ between 0.50 and 0.99. The viscosity of the *n*-paraffins at the boiling point may be taken to be 0.21 c.p. These values have been read from a large-scale chart and represent the mean values for the experimental data. It will be seen that data published in the literature for individual *n*-paraffins differs from these mean values; but it is believed that the difference is due to extreme difficulty in isolating

individual members free from all other isomers of the same molecular weight and of similar boiling characteristics. The first three members are not included, as their shape factor is not the same as that of the higher members⁽¹⁾.

While the utility of the equations developed here is of a certain degree of importance, as viscosity-temperature relationships may be determined from only two experimental values, the significance of the equations is of greater value and interest. These equations render certain tentative deductions of previous studies into quantitative statements.

TABLE II.

$\frac{T}{T_B}$	μ_P	$\frac{T}{T_B}$	μ_P	$\frac{T}{T_B}$	μ_P	$\frac{T}{T_B}$	μ_P	$\frac{T}{T_B}$	μ_P
0.50	4.4	0.60	1.47	0.70	0.685	0.80	0.405	0.90	0.278
0.51	3.9	0.61	1.34	0.71	0.643	0.81	0.388	0.91	0.269
0.52	3.45	0.62	1.23	0.72	0.605	0.82	0.371	0.92	0.260
0.53	3.04	0.63	1.14	0.73	0.572	0.83	0.357	0.93	0.251
0.54	2.70	0.64	1.04	0.74	0.542	0.84	0.342	0.94	0.245
0.55	2.43	0.65	0.970	0.75	0.513	0.85	0.330	0.95	0.239
0.56	2.18	0.66	0.895	0.76	0.488	0.86	0.318	0.96	0.233
0.57	1.96	0.67	0.835	0.77	0.463	0.87	0.306	0.97	0.228
0.58	1.78	0.68	0.779	0.78	0.441	0.88	0.296	0.98	0.223
0.59	1.62	0.69	0.730	0.79	0.422	0.89	0.287	0.99	0.218

Conclusions.

The first point to notice is the unification of the types of liquid, such as associated and unassociated, complex compounds and simple metallic liquids, etc., into one class which is governed by one law. The only pure liquid which is definitely known not to obey this law is liquid helium, the viscosity of which rises with rise in temperature. This fact has already been discussed⁽¹⁾. Otherwise, all the liquids listed, which were studied with sufficient detail to ensure that at least three or four determinations are found reliably, appear to be governed by the formulæ :—

$$\log \mu = K' - K_1' \left(\frac{T}{T_B} \right) + K_2' \left(\frac{T}{T_B} \right)^2 - K_3' \left(\frac{T}{T_B} \right)^3 + \dots;$$

$$\log \mu = K'' - K_1''(T) + K_2''(T)^2 - K_3''(T)^3 + \dots;$$

$$\log \mu = A - B \left[18.572 \left(\frac{T}{T_B} \right) - 17.505 \left(\frac{T}{T_B} \right)^2 + 5.900 \left(\frac{T}{T_B} \right)^3 + \dots \right];$$

$$\log \mu = A - B[6.283 - \log [\mu_P]];$$

$$\log \mu = C + B \log [\mu_P];$$

where

A, B, C, and variants of K are constants ;

μ = viscosity value at $T^\circ \text{K}$;

T_B = boiling point in $^\circ \text{K}$;

μ_P = viscosity of normal paraffins at same value of $\frac{T}{T_B}$ as that

of the liquid, the viscosity of which is μ .

Theories discussing the mechanism of viscosity of liquids will therefore be required to render the behaviour of all types of liquid governable by one equation, with no reservation as to degree of association, etc. Theories put forward up till now are generally based on the simple exponential formula :

$$\begin{aligned}\mu &= A e^{B/T} \\ &= A e^{Q/RT},\end{aligned}$$

where A and B are constants and $B = \frac{Q}{R}$, the ratio of energy of activation of viscosity to the gas constant.

Further, it is known that

$$\frac{L}{Q} = n, \text{ constant,}$$

where L = molal latent heat of vaporization.

The statement is often made that, whereas Q is a constant for normal paraffins and other unassociated liquids, it is variant with temperature for associated liquids. The formulæ developed show that Q is variant with temperature for all liquids studied here, and the difference between associated and unassociated liquids is entirely a matter of degree of the variation.

Thus

$$\begin{aligned}\log_e \left(\frac{\mu_1}{\mu_2} \right) &= \frac{1}{R} \left(\frac{Q_1}{T_1} - \frac{Q_2}{T_2} \right) \\ &= \frac{1}{2.308} [K_1''(T_2 - T_1) + K_2''(T_2^2 - T_1^2) + K_3''(T_2^3 - T_1^3) -].\end{aligned}$$

Further to unifying all liquids (excepting helium) in chemically pure state and of constant boiling point, there is indication that these formulæ tend to unify the laws of viscous flow independently of state of matter, *i. e.*, for gases, liquids and solids. A characteristic difference between gases and liquids has always been taken as the difference in their viscosity-temperature relationship-gases having an increase in their viscosity with

rise in temperature and liquids a decrease. From previous studies, it was tentatively concluded that a rise in temperature affected the viscosity coefficient of liquids in two essentially different ways: one effect being an increase in viscosity with temperature and the other and more generally known effect being a decrease. It is seen here that the increase in the logarithm of the viscosity of liquids with temperature is definite and substantial, being in the case of *n*-paraffins equal to $17.505 \left(\frac{T}{T_B} \right)^2$, as compared to the chief terms responsible for decrease of viscosity with temperature $\left[18.572 \left(\frac{T}{T_B} \right) + 5.900 \left(\frac{T}{T_B} \right)^3 \right]$. In all the other liquids tested by these charts the ratio of the increase to decrease in the logarithm of viscosity due to rise in temperature is of the same magnitude,

$$i. e., \quad \frac{\text{increase}}{\text{decrease}} = \frac{B \ 17.505 \left(\frac{T}{T_B} \right)^2}{B \left[18.572 \left(\frac{T}{T_B} \right) + 5.900 \left(\frac{T}{T_B} \right)^3 \right]}$$

$$= \frac{17.505 \left(\frac{T}{T_B} \right)}{\left[18.572 + 5.900 \left(\frac{T}{T_B} \right)^2 \right]}.$$

Again, theories explaining the nature of the viscosity of liquids do not take this fact into account, but rather assume the viscosity of liquids only to decrease with rise in temperature. It may be explained that the increase is due to a similar cause, which gives an increase in viscosity of gases, *i. e.*, a greater loss in energy due to imperfect elastic transference of momentum by collision. The decrease is due to drop in the potential barriers of Eyring, or similar explanations. The important point is that liquids and gases are behaving essentially alike in these respects. It was found⁽³⁾ that hydrocarbon gases under great pressures show a decrease in viscosity with rise in temperature. It is also significant that Sutherland's equation for gases, put in logarithmic form, is

$$\log \mu = K + \frac{3}{2} (\log T) - \log (T + c),$$

i. e., contains both rising and falling functions of temperatures.

The flow of liquids and solid crystals has already been correlated⁽⁴⁾. Thus it appears that in all probability a general equation exists, giving the change of viscosity of fluids with temperature independently of the state of the fluid. This fact imposes another criterion on any theory

attempting to explain the nature of viscosity. Not only must it be general to all liquids but it should be general to all fluids too, if it represents the true nature of viscosity. In other words, in both gases and liquids a rise in temperature results in (1) an increase in the energy of activation required for viscous flow due to collisions of the molecules, (2) a decrease in the viscous drag due to lower potential barriers between the molecules at the higher temperatures.

References.

- (1) Nissan, Clark and Nash, J. Inst. Pet. xxvi, pp. 154-222 (1940); cf. also Sc. Abstracts, No. 2044, xliii. (511), p. 523 (25 July, 1940).
- (2) Nissan, and Clark, 'Nature,' cxliii. p. 722 (1939).
- (3) Nelson, Oil Gas J. xxxviii. (9), p. 50 (13 July, 1939).
- (4) Ewell, J. Appl. Phys. ix, p. 252 (1938).

Department of Oil Engineering and Refining,
The University, Birmingham.

XLIV. *Steady Flow of a Viscous Fluid through a Leaky Tube.*

By F. J. TURTON, B.Sc., A.R.C.Sc., City and Guilds College *.

[Received September 1, 1941.]

§ 1. *Introduction.*

THE steady flow of a viscous liquid through straight tubes of uniform circular section and with impervious walls has been investigated experimentally by Poiseuille †, whose results are confirmed by hydrodynamics ‡. In biology, however, we meet with similar flow through tubes with membranous walls, such as the fine capillaries through the walls of which nutritious fluids get from the blood to working (*e. g.*, muscle) cells. There is experimental evidence § that, if p_i be the pressure at any point on the membrane, inside the capillary, and p_o the pressure outside at that point, then the rate of fluid flow per unit area through the membrane at the point is $k(p_i - p_o)$, k being a constant called the "leak constant"; p_i will, of course, vary from point to point along the capillary, but p_o is usually constant. In the hydrodynamical investigation which follows, this provides a boundary condition additional to the usual assumption that there is no tangential motion, relative to the wall, of the fluid immediately in contact with it.

For steady viscous flow through non-leaky tubes the flux is $\pi a^4(p_1 - p_2)/8\eta l$, where a is the radius, l the length, p_1 and p_2 the pressures at the ends and η the coefficient of viscosity; and the pressure gradient along the tube is constant and equal to $-(p_1 - p_2)/l$.

With leaky walls, the following investigation shows that the pressure gradient may approximate closely to this constancy, but for certain conditions known to exist in practice, the mathematical results show considerable departure from constant pressure gradient.

The following range of values has been considered, as likely to include all cases which can be studied in practice:—

a from 7×10^{-4} to 3×10^{-4} cms.

k from 6×10^{-11} to 6×10^{-9} ccs./sec. per cm.² per dyne/cm.² or cm./sec. per dyne/cm.²

* Communicated by Dr. W. G. Bickley.

† *Comptes Rendus*, tt. xi., xii. (1840–41); *Mém. des Sav. Étrangers*, t. ix. (1846).

‡ Lamb, 'Hydrodynamics,' 3rd ed., p. 543.

§ Landis, *Amer. J. of Physiology*, lxxxii. p. 217 (1927).

l from 5×10^{-3} to 2×10^{-1} cm.

η from 1×10^{-2} to 6×10^{-2} poise.

$(p_1 - p_0)$ up to 2×10^4 dynes/cm.² and $(p_2 - p_0)$ from -2×10^4 to 2×10^4 dynes/cm.² (It is assumed that $p_1 \geq p_2$.)

Inertia terms in the equations have been omitted at an early stage, the equations otherwise being intractable. The writer has examined the ratio of terms omitted to terms retained in a number of cases, and believes the inertia terms to be negligible for the range of values just quoted.

The manipulative operations on Bessel functions have in the main been omitted. The necessary equations will be found in Gray, Mathews and MacRobert, 'Bessel Functions'; McLachlan, 'Bessel Functions for Engineers,' etc.

§2. Fundamental Equations.

If u, v, w are the component velocities of that element of fluid at (x, y, z) at time t and p is the pressure, the equations of steady motion of a viscous incompressible fluid, having coefficient of viscosity η and density ρ , are

$$\begin{aligned} \rho \left(u \frac{\partial u}{\partial x} + v \frac{\partial u}{\partial y} + w \frac{\partial u}{\partial z} \right) + \frac{\partial p}{\partial x} &= \eta \left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} \right), \\ \rho \left(u \frac{\partial v}{\partial x} + v \frac{\partial v}{\partial y} + w \frac{\partial v}{\partial z} \right) + \frac{\partial p}{\partial y} &= \eta \left(\frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} + \frac{\partial^2 v}{\partial z^2} \right), \\ \rho \left(u \frac{\partial w}{\partial x} + v \frac{\partial w}{\partial y} + w \frac{\partial w}{\partial z} \right) + \frac{\partial p}{\partial z} &= \eta \left(\frac{\partial^2 w}{\partial x^2} + \frac{\partial^2 w}{\partial y^2} + \frac{\partial^2 w}{\partial z^2} \right), \\ \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} &= 0^*. \end{aligned}$$

If the inertia terms be omitted, by differentiation we obtain

$$\frac{\partial^2 p}{\partial x^2} + \frac{\partial^2 p}{\partial y^2} + \frac{\partial^2 p}{\partial z^2} = 0^\dagger,$$

and thence the third equation above gives

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right)^2 w = 0. \quad \dots \dots \dots (1)$$

For the present problem, in which we assume radial symmetry, change to cylindrical polar co-ordinates (r, θ, z) by $x = r \cos \theta$, $y = r \sin \theta$. Denote by s the radial component of the velocity; the tangential component will be identically zero, and p, w and s will be independent of θ .

* Lamb, *loc. cit.* p. 538.

† Lamb, *loc. cit.* p. 550.

The equations become *

$$\frac{1}{\eta} \frac{\partial p}{\partial r} + \frac{\rho}{\eta} \left(s \frac{\partial s}{\partial r} + w \frac{\partial s}{\partial z} \right) = \frac{\partial^2 s}{\partial r^2} + \frac{1}{r} \frac{\partial s}{\partial r} + \frac{\partial^2 s}{\partial z^2} - \frac{s}{r^2}, \quad (2)$$

$$\frac{1}{\eta} \frac{\partial p}{\partial z} + \frac{\rho}{\eta} \left(s \frac{\partial w}{\partial r} + w \frac{\partial w}{\partial z} \right) = \frac{\partial^2 w}{\partial r^2} + \frac{1}{r} \frac{\partial w}{\partial r} + \frac{\partial^2 w}{\partial z^2}, \quad (3)$$

$$\frac{\partial s}{\partial r} + \frac{s}{r} + \frac{\partial w}{\partial z} = 0, \quad (4)$$

or, omitting the inertia terms,

$$\frac{1}{\eta} \frac{\partial p}{\partial r} = \frac{\partial^2 s}{\partial z^2} + \frac{1}{r} \frac{\partial s}{\partial r} + \frac{\partial^2 s}{\partial r^2} - \frac{s}{r^2}, \quad (5)$$

$$\frac{1}{\eta} \frac{\partial p}{\partial z} = \frac{\partial^2 w}{\partial r^2} + \frac{1}{r} \frac{\partial w}{\partial r} + \frac{\partial^2 w}{\partial z^2}, \quad (6)$$

$$\frac{\partial s}{\partial r} + \frac{s}{r} + \frac{\partial w}{\partial z} = 0 \quad \text{or} \quad \frac{1}{r} \frac{\partial(rs)}{\partial r} + \frac{\partial w}{\partial z} = 0. \quad (7)$$

Differentiating the last partially with respect to r , we have

$$\frac{\partial^2 s}{\partial r^2} + \frac{1}{r} \frac{\partial s}{\partial r} - \frac{s}{r^2} + \frac{\partial^2 w}{\partial r \partial z} = 0;$$

hence, from (5),

$$\frac{1}{\eta} \frac{\partial p}{\partial r} = - \frac{\partial^2 w}{\partial r \partial z} + \frac{\partial^2 s}{\partial z^2}. \quad (5a)$$

Also, in polar co-ordinates with radial symmetry, (1) becomes

$$\left(\frac{\partial^2}{\partial r^2} + \frac{1}{r} \frac{\partial}{\partial r} + \frac{\partial^2}{\partial z^2} \right) w = 0. \quad (1a)$$

§ 3. Solution of the Equations.

$$\left(\frac{\partial^2}{\partial r^2} + \frac{1}{r} \frac{\partial}{\partial r} + \frac{\partial^2}{\partial z^2} \right) w = 0. \quad (1a)$$

Try $w = e^{\alpha z}$. R , R being a function of r only. This is a solution if

$$\left(\frac{d^2}{dr^2} + \frac{1}{r} \frac{d}{dr} + \alpha^2 \right) R = 0.$$

$$\text{Now} \quad \left(\frac{d^2}{dr^2} + \frac{1}{r} \frac{d}{dr} + \alpha^2 \right) \left(\frac{d^2}{dr^2} + \frac{1}{r} \frac{d}{dr} + \beta^2 \right) R = 0, \quad \beta \neq \alpha,$$

where R is to be finite when $r=0$, is satisfied by

$$R = C_1 J_0(\alpha r) + C_2 J_0(\beta r),$$

C_1 and C_2 being arbitrary.

* Cf. Bassett, 'Hydrodynamics,' ii. p. 244.

When $\beta \rightarrow \alpha$, a second solution will be

$$\frac{\partial}{\partial \alpha} \{C_2 J_0(\alpha r)\} \quad \text{or} \quad -C_2 r J_1(\alpha r).$$

$$w = P e^{\alpha z} J_0(\alpha r) + Q e^{\alpha z} \cdot \alpha r \cdot J_1(\alpha r)$$

thus satisfies (1a) for any α .

Substituting this value in (7) gives

$$\frac{1}{r} \frac{\partial(rs)}{\partial r} = -\alpha P e^{\alpha z} J_0(\alpha r) - Q e^{\alpha z} \cdot \alpha^2 r \cdot J_1(\alpha r),$$

whence $rs = -P e^{\alpha z} \cdot r \cdot J_1(\alpha r) - Q e^{\alpha z} \cdot \alpha r^2 \cdot J_2(\alpha r) + \phi(z)$

and $s = -P e^{\alpha z} J_1(\alpha r) - Q e^{\alpha z} \cdot \alpha r \cdot J_2(\alpha r),$

for, since the velocity s must be finite at the axis, we must have $\phi(z) \equiv 0$.

Substituting in (5a),

$$\frac{1}{\eta} \frac{\partial p}{\partial r} = -2Q e^{\alpha z} \cdot \alpha^2 \cdot J_1(\alpha r),$$

giving

$$p = \varpi(z) + 2Q\eta\alpha e^{\alpha z} J_0(\alpha r).$$

Substituting in (6),

$$\frac{1}{\eta} \frac{\partial p}{\partial z} = 2Q\alpha^2 e^{\alpha z} J_0(\alpha r),$$

whence

$$p = \theta(r) + 2Q\eta\alpha e^{\alpha z} J_0(\alpha r).$$

Compare

$$p = \varpi(z) + 2Q\eta\alpha e^{\alpha z} J_0(\alpha r).$$

The results agree if $\theta(r) = \varpi(z) = c$, a constant.

Thus, C being a constant, the equations are satisfied by

$$p = C + \Sigma 2Q\eta\alpha e^{\alpha z} J_0(\alpha r), \quad \dots \dots \dots (8)$$

$$w = \Sigma \{P e^{\alpha z} J_0(\alpha r) + Q e^{\alpha z} \cdot \alpha r \cdot J_1(\alpha r)\}, \quad \dots \dots \dots (9)$$

$$s = \Sigma \{-P e^{\alpha z} J_1(\alpha r) - Q e^{\alpha z} \cdot \alpha r \cdot J_2(\alpha r)\}. \quad \dots \dots \dots (10)$$

For viscous flow in a tube of radius a we must have $w=0$ when $r=a$ for all z and every α , P , Q .

This will be so if

$$P J_0(\alpha a) + Q \alpha a J_1(\alpha a) = 0 \text{ for every } \alpha, P, Q. \quad \dots \dots \dots (11)$$

We must also have $s=k(p-p_0)$ when $r=a$ for all z and every α , P , Q , i. e.,

$$\Sigma \{-P e^{\alpha z} J_1(\alpha a) - Q e^{\alpha z} \cdot \alpha a \cdot J_2(\alpha a)\} \equiv k\{C + \Sigma 2Q\eta\alpha e^{\alpha z} J_0(\alpha a) - p_0\}.$$

This will be satisfied if

$$C = p_0, \quad \dots \dots \dots (12)$$

and if

$$-P J_1(\alpha a) - Q \alpha a J_2(\alpha a) = 2kQ\eta\alpha J_0(\alpha a) \text{ for every } \alpha, P, Q. \quad (13)$$

Eliminating P and Q between (11) and (13)

$$\frac{J_1^2(\alpha a) - J_0(\alpha a)J_2(\alpha a)}{J_0^2(\alpha a)} = \frac{2k\eta}{a} \quad (14)$$

This equation is even in αa and so has pairs of roots of equal magnitude and opposite sign. The full discussion of the equation is left until later; here we state that if $2k\eta/a < 0.9397 \dots$ there are only two real roots, approximating to $\pm \sqrt{16k\eta/a}$ when small.

In the biological problem, $2k\eta/a$ is very much less than $0.9397 \dots$; in fact, less than 5×10^{-5} for any of the values in the introduction. For such small values $\alpha^2 a^2$ approximates to $16k\eta/a$ within 0.013 per cent., and is less than 4×10^{-4} . Consequently, $J_0(\alpha r) = 1 - \alpha^2 r^2/2^2 + \dots$ approximates to 1 within less than 10^{-4} .

When calculating the coefficients for the two roots, α and $-\alpha$ say, we are able therefore to neglect the pressure variation across the tube and put $p = p_1$ and p_2 for two specified values of z , which it will be convenient to take as $z=0$ and $z=l$ respectively.

Restricting the summations to two roots, α and $-\alpha$, using suffixes $_1$ and $_2$ respectively for the coefficients (obtaining the second set by writing $-\alpha$ for α in the first) and using the relation (11) to get rid of P_1 and P_2 , we have, from (8), (9) and (10), with (12)

$$p = p_0 + 2Q_1\eta\alpha e^{\alpha z}J_0(\alpha r) - 2Q_2\eta\alpha e^{-\alpha z}J_0(\alpha r), \quad (8a)$$

$$w = (Q_1e^{\alpha z} + Q_2e^{-\alpha z})\{\alpha rJ_1(\alpha r) - \alpha aJ_1(\alpha a)J_0(\alpha r)/J_0(\alpha a)\}, \quad (9a)$$

$$s = (Q_1e^{\alpha z} - Q_2e^{-\alpha z})\{-\alpha rJ_2(\alpha r) + \alpha aJ_1(\alpha a)J_1(\alpha r)/J_0(\alpha a)\}. \quad (10a)$$

Put $2Q_1 = A - Be^{-\alpha l}$ and $2Q_2 = A - Be^{\alpha l}$, obtaining

$$p = p_0 + 2\eta\alpha\{A \sinh \alpha z + B \sinh \alpha(l-z)\}J_0(\alpha r), \quad (8b)$$

$$w = \{B \cosh \alpha(l-z) - A \cosh \alpha z\}\{\alpha aJ_1(\alpha a)J_0(\alpha r)/J_0(\alpha a) - \alpha rJ_1(\alpha r)\}, \quad (9b)$$

$$s = \{A \sinh \alpha z + B \sinh \alpha(l-z)\}\{-\alpha rJ_2(\alpha r) + \alpha aJ_1(\alpha a)J_1(\alpha r)/J_0(\alpha a)\}. \quad (10b)$$

Now, assuming $p = p_1$ at $z=0$ and $p = p_2$ at $z=l$, on the axis $r=0$,

$$p_1 - p_0 = 2\eta\alpha B \sinh \alpha l,$$

$$p_2 - p_0 = 2\eta\alpha A \sinh \alpha l, \text{ giving}$$

$$\begin{aligned} p - p_0 &= \{(p_1 - p_0) \sinh \alpha(l-z) + (p_2 - p_0) \sinh \alpha z\}J_0(\alpha r)/\sinh \alpha l \\ &= \{(p_1 - p_0) \sinh \alpha(l-z) + (p_2 - p_0) \sinh \alpha z\}/\sinh \alpha l, \text{ nearly,} \end{aligned}$$

$$\text{since } J_0(\alpha r) \doteq 1. \quad (15)$$

(Compare Poiseuille :

$$p - p_0 = (p_1 - p_0) - (p_1 - p_2)z/l$$

$$= \{(p_1 - p_0)\alpha(l-z) + (p_2 - p_0)\alpha z\}/\alpha l \text{ for any } \alpha).$$

$$w = \frac{(p_1 - p_0) \cosh \alpha(l-z) - (p_2 - p_0) \cosh \alpha z}{2\eta\alpha \sinh \alpha l} \left\{ \frac{\alpha a J_1(\alpha a) J_0(\alpha r)}{J_0(\alpha a)} - \alpha r J_1(\alpha r) \right\}$$

$$= \{(p_1 - p_0) \cosh \alpha(l-z) - (p_2 - p_0) \cosh \alpha z\} \alpha(a^2 - r^2)/4\eta \sinh \alpha l, \quad (16)$$

nearly, αr being so small.

$$s = \frac{(p_1 - p_0) \sinh \alpha(l-z) + (p_2 - p_0) \sinh \alpha z}{2\eta\alpha \sinh \alpha l} \left\{ \frac{\alpha a J_1(\alpha a) J_1(\alpha r)}{J_0(\alpha a)} - \alpha r J_2(\alpha r) \right\}$$

$$= \{(p_1 - p_0) \sinh \alpha(l-z) + (p_2 - p_0) \sinh \alpha z\} \alpha^2 r(2a^2 - r^2)/16\eta \sinh \alpha l \quad (17)$$

nearly, αr being so small.

$$= (p - p_0) \alpha^2 r(2a^2 - r^2)/16\eta \quad (17a)$$

$$= (p - p_0) k r(2a^2 - r^2)/a^3 \text{ if also } \alpha^2 = 16k\eta/a^3. \quad (17b)$$

The flux in at $z=0$ is $\int_0^a w_0 \cdot 2\pi r dr$. Substituting for w with $z=0$, and integrating, we get

$$\{(p_1 - p_0) \cosh \alpha l - (p_2 - p_0)\} 2\pi k a J_0(\alpha a)/\alpha \sinh \alpha l$$

$$= \{(p_1 - p_0) \cosh \alpha l - (p_2 - p_0)\} 2\pi k a/\alpha \sinh \alpha l, \text{ nearly.} \quad (18)$$

The flux out at $z=l$ is similarly

$$\{(p_1 - p_0) - (p_2 - p_0) \cosh \alpha l\} 2\pi k a/\alpha \sinh \alpha l, \text{ nearly.} \quad (19)$$

The total leakage is therefore

$$(p_1 + p_2 - 2p_0)(\cosh \alpha l - 1) 2\pi k a/\alpha \sinh \alpha l, \text{ nearly.} \quad (20)$$

It is readily verified that this is equal to

$$\int_0^l 2\pi k a (p - p_0) dz.$$

§ 4. Approximate Values when αl is small.

The least value of α for the values of k , η and a in the introduction is about 0.16, and l ranges from about 5×10^{-3} upwards, so it is evident that αl will often be small.

(a) When $\alpha^4 l^4/4!$ is negligible compared with 1, $\cosh \alpha l$ may be replaced by $1 + \alpha^2 l^2/2$ and $\sinh \alpha l$ by $\alpha l + \alpha^3 l^3/6$. Also $\alpha^2 = 16k\eta/a^3$.

$$\text{Then inflow} = \{(p_1 - p_0)(1 + \alpha^2 l^2/2) - (p_2 - p_0)\} 2\pi k a/\alpha^2 l(1 + \alpha^2 l^2/6)$$

$$= \{\pi a^4(p_1 - p_2)/8\eta l\} + \{(2p_1 + p_2 - 3p_0)\pi k a l/3\}, \text{ nearly.}$$

The outflow is similarly

$$\{\pi a^4(p_1 - p_2)/8\eta l\} - \{(p_1 + 2p_2 - 3p_0)\pi k a l/3\}, \text{ nearly,}$$

and the leakage $(p_1 + p_2 - 2p_0)\pi k a l$, nearly.

Similarly,

$$p-p_0=\{(p_1-p_0)(1-z/l)+(p_2-p_0)(z/l)\}-8k\eta z(l-z) \\ \times \{l(2p_1+p_2-3p_0)-z(p_1-p_2)\}/3la^3.$$

(b) When $\alpha^2 l^2/2$ is negligible compared with 1, these simplify still further.

$$\text{Inflow}=\text{outflow}=\pi a^4(p_1-p_2)/8\eta l,$$

$$p-p_0=(p_1-p_0)(1-z/l)+(p_2-p_0)(z/l),$$

i. e., Poiseuille's results for non-leaky tubes; the ratio leak/flow is negligible.

§ 5. Roots of $\{J_1^2(x)-J_0(x) \cdot J_2(x)\}/J_0^2(x)=2k\eta/a$.

(a) The roots may be found by drawing the horizontal straight line $y_1=2k\eta/a$ on the graph of

$$y_2=\{J_1^2(x)-J_0(x) \cdot J_2(x)\}/J_0^2(x).$$

$J_0(x)$ and $J_2(x)$ are even functions of x , so that if $x=\alpha a$ is a root, then $x=-\alpha a$ is also a root. Thus the discussion of the graph of y_2 may be confined to positive x .

$J_0(x)$, $J_1(x)$ and $J_2(x)$, and consequently y_2 , are real and single-valued for every real value of x .

Writing $y_3=J_1^2(x)-J_0(x) \cdot J_2(x)$, $y_3=0$ when $x=0$. It turns where $J_0(x)=0$ or $J_2(x)=0$; where $J_0(x)=0$, its second derivative is $-J_2^2(x)$, hence here y_3 has maximum values, the values being those of $J_1^2(x)$ at these points; where $J_2(x)=0$, its second derivative is $-J_0^2(x)$, hence here y_3 has minimum values, the values being those of $J_1^2(x)$ at these points. Also y_3 is finite and continuous for all real x . From these facts it follows that y_3 is everywhere positive (except at $x=0$).

Hence also y_2 is everywhere positive (except at $x=0$). Replacing $J_2(x)$ by $(2J_1(x)/x)-J_0(x)$, we find $y_2=1$ where $J_1(x)=0$ (viz., at $x=3.832, 7.016, 10.173, 13.323$, etc.), and where $xJ_1(x)=2J_0(x)$ (viz., at $x=4.291, 7.288, 10.366, 13.472$, etc.). These approximate, one from each group, to $n\pi+\pi/4$, which has values $3.927, 7.069, 10.210, 13.352$, etc.

$J_0(x)$ vanishes at $x=2.405, 5.52, 8.65, 11.79$, etc., which approximate to $n\pi+3\pi/4$; the values of the latter are $2.36, 5.50, 8.64, 11.78$, etc. so the approximation is good, even for the first. Hence y_2 has asymptotes $x=2.405, 5.52$, etc., approaching its asymptotes on both sides with y_2 positive.

The asymptotic expansions for the Bessel functions give, to order $1/x$,

$$y_2 \doteq 2(1+\cos 2x/2x)/(1+\sin 2x-\cos 2x/4x)=y_4 \text{ say.}$$

Now y_4 has values $2(1+3/4x)$ at $x=n\pi$, 1 at $x=n\pi+\pi/4$, $2(1-3/4x)$ at $x=n\pi+\pi/2$ and $\rightarrow\infty$ as $x\rightarrow n\pi+3\pi/4$. Further, y_4 approximates to $2/(1+\sin 2x)=\sec(x-\pi/4)$ when x is large, the approximation being particularly good when x is nearly $n\pi+\pi/4$, for there $\cos 2x$ is zero. Hence, except near $x=0$, y_4 (and therefore y_2) approximates to a succession of festoons with asymptotes near $x=n\pi+3\pi/4$ and minima near $x=n\pi+\pi/4$, $y=1$. From $x=0$ to the first asymptote, $x=2.405$, y_2 rises steadily, giving a half-festoon.

(b) Turning values of

$$y_2 = \{J_1^2(x) - J_0(x) \cdot J_2(x)\} / J_0^2(x).$$

If y_2 turns at $x=X$, write briefly J_0 for $J_0(X)$, etc. Then y_2 turns where

$$J_0(2J_1J_1' - J_0J_2' - J_2J_0') = 2J_0'(J_1^2 - J_0J_2)$$

$$J_0\{2J_1(J_0 - J_1/X) - J_0(J_1 - 2J_2/X) + J_1J_2\} = -2J_1(J_1^2 - J_0J_2)$$

$$J_0\{XJ_1(J_0 + J_2) - 2J_1^2 + 2J_0J_2\}/X = -2J_1(J_1^2 - J_0J_2),$$

or

$$2J_0^2J_2/X = -2J_1(J_1^2 - J_0J_2),$$

i. e., where

$$(J_1^2 - J_0J_2)/J_0^2 = (y_2)_{\text{minimum}} = -J_2/XJ_1.$$

If we put $J_1/J_0 = X^{-1} - \epsilon$ in this, after replacing J_2 by $(2J_1/X) - J_0$, and rearrange, we have

$$X\epsilon = (1 + X^2\epsilon^2)/(1 + X^2\epsilon^2 + X^2), \quad (21)$$

hence

$$0 < X\epsilon < 1.$$

$(y_2)_{\text{min.}}$ in terms of X and ϵ becomes $1 + \epsilon^2 - X^{-2}$, hence exceeds $1 - X^{-2}$; it may also be written $1 - (1 - X^2\epsilon^2)/X^2$ which is less than 1, since $X\epsilon < 1$. Hence $1 > (y_2)_{\text{min.}} > 1 - X^{-2}$.

Further, (21) may be rearranged as $X\epsilon = \{1 - X\epsilon(1 - X\epsilon) - X^3\epsilon^3\}/X^2$, hence $X\epsilon < X^{-2}$ and $\epsilon^2 < X^{-6}$. Since $(y_2)_{\text{min.}} = 1 + \epsilon^2 - X^{-2}$, it is therefore less than $1 - X^{-2} + X^{-6}$, so that

$$1 - X^{-2} + X^{-6} > (y_2)_{\text{min.}} > 1 - X^{-2}. \quad (22)$$

By putting $X\epsilon = (1 - \delta)X^{-2}$, it may similarly be shown (using also $X^2 > (2.40)^2$) that $\delta < X^{-2}$, and thence that

$$1 - X^{-2} + X^{-6} > (y_2)_{\text{min.}} > 1 - X^{-2} + X^{-6} - 2X^{-8}.$$

From (22) we can show that minimum values of y_2 form a monotonously increasing sequence. If X_s and X_{s+1} be the values of x at successive minima, it is sufficient to show that

$$1 - X_{s+1}^{-2} > 1 - X_s^{-2} + X_s^{-6},$$

for the minimum y_2 at X_{s+1} exceeds the left-hand side, whilst the right-hand side exceeds the minimum at X_s ; i. e., it is sufficient to show that

$$(X_{s+1}^2 - X_s^2)/X_s^2 X_{s+1}^2 > X_s^{-6} \text{ or that } X_{s+1} - X_s > X_{s+1}^2/(X_{s+1} + X_s)X_s^4.$$

Now successive values of X lie between 3.832 and 4.291, 7.016 and 7.288, 10.173 and 10.366, 13.323 and 13.472, etc., at which $y_2=1$; hence $X_{s+1}-X_s$ is certainly less than, say, 4 and greater than, say, 2.5. Also $X_s > 3.832$. Thus

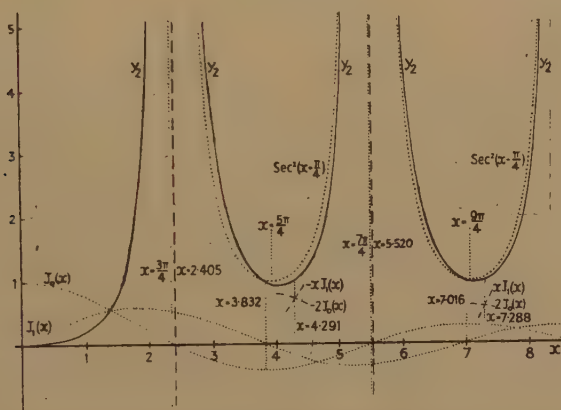
$$\begin{aligned} X_{s+1}^2/(X_{s+1}+X_s)X_s^4 &< (X_s+4)^2 \cdot 2X_s^{-5} \\ &< \frac{1}{2}(X_s^{-3}+8X_s^{-4}+16X_s^{-5}) < 0.1 \text{ say,} \end{aligned}$$

whilst $X_{s+1}-X_s > 2.5$.

Thus $X_{s+1}-X_s$ is much greater than $X_{s+1}^2/(X_{s+1}+X_s)X_s^4$. Hence the minima for y_2 form a monotonously increasing sequence.

The first minimum beyond $x=0$ has been found by superimposing $y_5=-J_2(x)/xJ_1(x)$ on y_2 in a region around the point (fig. 3). The

Fig. 1.



$y_2 = \{J_1^2(x) - J_0(x) \cdot J_2(x)\} / J_0^2(x)$ for positive x to $x=8.5$, thus—

intersection is almost orthogonal at $x=4.066 \dots$, the minimum value for y_2 there being $0.9397 \dots$; at $X=4.066$, $1-X^{-2}+X^{-6}=0.93973$, whilst $1-X^{-2}+X^{-6}-2X^{-8}=0.93970$, confirming the value $0.9397 \dots$ found there. Hence successive minima at values of x approximating to $n\pi + \pi/4$ rise steadily from $0.9397 \dots$ to 1.

Graphs of y_2 and associated functions are given in figs. 1, 2 and 3.

(c) Thus if $2k\eta/a < 0.9397 \dots$, the equation

$$\{J_1^2(x) - J_0(x) \cdot J_2(x)\} / J_0^2(x) = 2k\eta/a$$

has but one positive real root; as $2k\eta/a$ rises from $0.9397 \dots$ to 1, there are 3, 5, 7, etc., positive real roots; when $2k\eta/a$ exceeds 1, there is an infinite number of real positive roots; and there are negative roots, in each case, of the same magnitude.

Fig. 2.

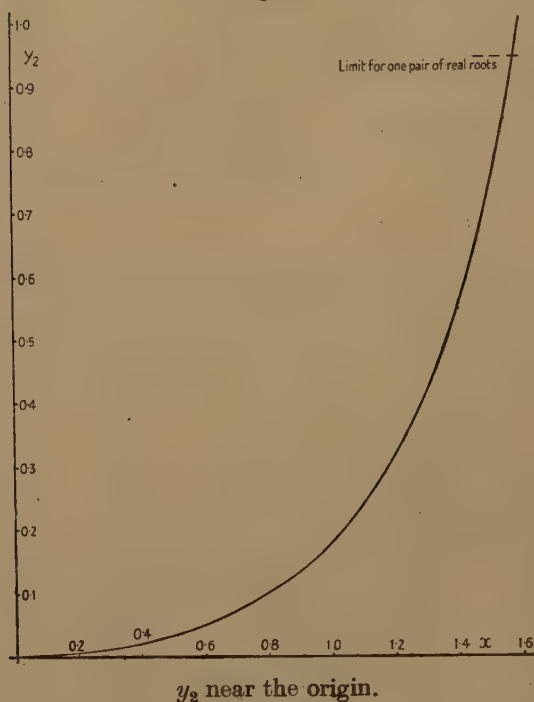
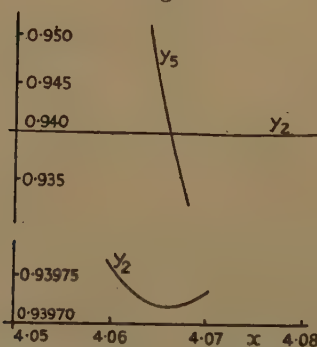


Fig. 3.

 y_2 and y_5 near the first minimum of y_2 .

In the biological problem we are concerned only with small values of $2k\eta/a$. Now

$$\begin{aligned} \{J_1^2(x) - J_0(x) \cdot J_2(x)\} / J_0^2(x) &= \{(x^2/4)(1 - x^2/8)^2 - (x^2/8)(1 - x^2/12) \\ &\quad \times (1 - x^2/4)\} / (1 - x^2/4)^2, \\ &= (x^2/4)/2 + 2(x^2/4)^2/3. \end{aligned}$$

Thus equation (14) becomes $z/2 + 2z^2/3 = 2k\eta/a$ where $z = \alpha^2 a^2/4$. Iterating this equation,

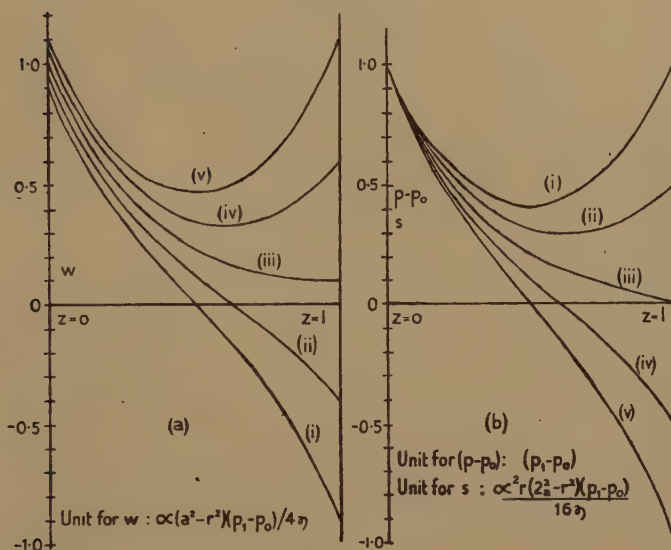
$$\alpha^2 a^2 = (16k\eta/a)(1 - 16k\eta/3a).$$

For all the values in the introduction, the second factor on the right is between 0.999872 and 1. Hence

$$\alpha^2 a^2 = 16k\eta/a \text{ is sufficiently accurate,}$$

or
$$\alpha^2 = 16k\eta/a^3.$$

Fig. 4.



Graph.	(i)	(ii)	(iii)	(iv)	(v)
$(p_2 - p_0)/(p_1 - p_0)$	1	0.5	0	-0.5	-1

Variation of (a) w and of (b) $(p-p_0)$ and s , along the tube when $al=3$, for five values of $(p_2-p_0)/(p_1-p_0)$.

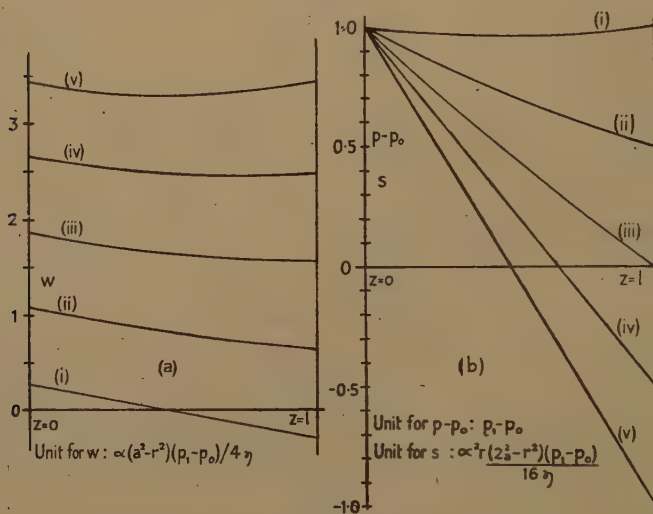
This approximation for α^2 depends on neglecting $16k\eta/3a$ compared with 1. The variation of pressure across the tube depends on $\alpha^2 r^2/4$ compared with 1 (when small) and $\alpha^2 r^2/4$ varies from 0 to $\alpha^2 a^2/4$, i. e., from 0 up to about $4k\eta/a$; so it is evident that in critical cases where the pressure variation can only just be neglected, the second factor in the expression above for $\alpha^2 a^2$ will have to be included. Such values are, however, at present beyond known or probable experimental values.

§ 6. Graphs.

Two sets of graphs, for $\alpha l = 3$ and for $\alpha l = 0.6$, are given. The first roughly corresponds to those values of the introduction which give greatest departure from the Poiseuille pressure graph; the second is more likely to be typical. In each set, one graph shows the variation of w with z , for various values of $(p_2 - p_0)/(p_1 - p_0)$; the other the variation of $(p - p_0)$ and of s in the same way. (Figs. 4 (a), 4 (b), 5 (a) and 5 (b).)

Further graphs show the variation of pressure along the tube for various values of αl . This variation can also be interpreted as for various

Fig. 5.



Graph.	(i)	(ii)	(iii)	(iv)	(v)
$(p_2 - p_0)/(p_1 - p_0)$	1	0.5	0	-0.5	-1

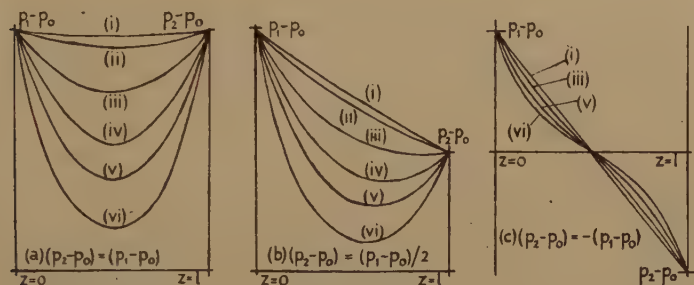
Variation of (a) w and of (b) $(p - p_0)$ and s , along the tube when $\alpha l = 0.6$, for five values of $(p_2 - p_0)/(p_1 - p_0)$.

values of the radius, other variables remaining constant; this is of particular interest to the biologist. Three typical values of $(p_2 - p_0)/(p_1 - p_0)$ are illustrated, viz. 1, $1/2$ and -1 . (Figs. 6 (a), 6 (b), 6 (c).)

Pressure variation across the tube can be neglected for values of α and of αl very much greater than those given by the values of the introduction. Thus for 1 per cent. pressure variation across the tube α can exceed 150 and αl may be 30 or even more. For such values of αl the

graph of the pressure along the tube, calculated from the equations of the present paper, shows a rapid fall at each end (when $p_1 - p_0$ and $p_2 - p_0$ are both positive, this is very marked; the graph approximates to $\lfloor \_ \_ \rfloor$; with $p_1 - p_0$ and $p_2 - p_0$ of opposite sign the graph approximates to $\lfloor \_ \_ \rfloor$, but much less closely for the same value of αl). Thus the main variation in pressure (so calculated) is almost all at the ends of the tube. This is in any case a place of great uncertainty; when there is no leak the stream-lines are not parallel in these regions and, in fact, in applying

Fig. 6.



Graph.	(i)	(ii)	(iii)	(iv)	(v)	(vi)
Values of αl	0.4	0.8	1.6	2.4	3.2	4.8
Relative magnitudes of radii ..	5.23	3.30	2.08	1.59	1.31	1

Variation of $(p - p_0)$ along the tube with αl or with radius,
for three values of $(p_2 - p_0)/(p_1 - p_0)$.

Poiseuille's equations the tube should be so long that this part can be ignored. The same difficulty arises here; and when practically all the variation of pressure occurs in these regions, it would appear unwise to attach any value to the results as here found; possibly also the inertia terms would not be negligible for such values.

§7. Summary.

In §1 the symbols a , k , l , η , p_0 , p_1 and p_2 are defined and the range of values considered is given.

In §2 p , s and w are defined and the equations obtained. Their solution is found in §3, and in §4 the approximate solutions (i) and (ii) below are obtained.

For all the values quoted in § 1, α as required below can be found with sufficient accuracy from $\alpha = \sqrt{16k\eta/a^3}$, found in § 5.

Then

- (i) If $\alpha^2 l^2/2$ is negligible in comparison with 1, Poiseuille's results apply :

$$\text{Inflow} = \text{outflow} = \pi a^4 (p_1 - p_2) / 8\eta l.$$

$$p = \{p_1(l-z) + p_2 z\} / l$$

and the (negligible) leakage is $(p_1 + p_2 - 2p_0)\pi k a l$.

- (ii) If $\alpha^4 l^4/4!$ is negligible in comparison with 1, but not $\alpha^2 l^2/2$:

$$\text{Inflow} = \{\pi a^4 (p_1 - p_2) / 8\eta l\} + (2p_1 + p_2 - 3p_0)\pi k a l / 3.$$

$$\text{Outflow} = \{\pi a^4 (p_1 - p_2) / 8\eta l\} - (p_1 + 2p_2 - 3p_0)\pi k a l / 3.$$

$$\text{Leakage} = (p_1 + p_2 - 2p_0)\pi k a l.$$

$$p - p_0 = (p_1 - p_0)(1 - z/l) + (p_2 - p_0)(z/l)$$

$$- 8k\eta z(l-z)\{2p_1 + p_2 - 3p_0 - (p_1 - p_2)z/l\} / 3a^3.$$

- (iii) Generally,

$$\text{Inflow} = \{(p_1 - p_0) \cosh \alpha l - (p_2 - p_0)\} 2\pi k a / \alpha \sinh \alpha l.$$

$$\text{Outflow} = \{(p_1 - p_0) - (p_2 - p_0) \cosh \alpha l\} 2\pi k a / \alpha \sinh \alpha l.$$

$$\text{Leakage} = (p_1 + p_2 - 2p_0)(\cosh \alpha l - 1) 2\pi k a / \alpha \sinh \alpha l.$$

$$p - p_0 = \{(p_1 - p_0) \sinh \alpha(l-z) + (p_2 - p_0) \sinh \alpha z\} / \sinh \alpha l.$$

$$w = \{(p_1 - p_0) \cosh \alpha(l-z) - (p_2 - p_0) \cosh \alpha z\} \alpha (a^2 - r^2) / 4\eta \sinh \alpha l.$$

$$s = kr(2a^2 - r^2)(p - p_0) / a^3.$$

- (iv) The above results hold only if $\alpha^2 a^2$ is negligible in comparison with 1 ; and α is found as above only if $16k\eta/3a$ is negligible in comparison with 1 ; this is the case for all the values quoted in § 1.

§ 8. Acknowledgements.

The preceding investigation arose from an enquiry by Dr. J. F. Danielli, of the Biochemical Laboratory, Cambridge, to whom the writer is indebted for the biological details and values and for much useful comment from a biologist's point of view.

The author would also like to thank Dr. W. G. Bickley for reading the manuscript, which resulted in considerable clarification and condensation.

The writer will be pleased to assist any biologist desiring to utilise these results ; a number of off-prints will be available.

XLV. *An Investigation concerning G. I. Taylor's Correlation Coefficient of Turbulence.*

By T. E. W. SCHUMANN, M.Sc., Ph.D.

[Received April 30, 1941.]

Introductory.

IN dealing with the theory of turbulence, Taylor (1920)† introduced the important concept of the function R_ξ , which he defined as the correlation coefficient between the turbulent velocity at a given point and the velocity at the same point after an interval of time ξ .

Since R_ξ is obviously a function of the time, it may conveniently be referred to as the "time correlation coefficient" of any continuously varying quantity, for it is applicable not only to turbulent velocities, but also to varying quantities such as temperature, pressure, etc.

The present paper is concerned with the derivation of an analytic expression for R_ξ . At the same time it is proposed to demonstrate that Taylor's deduction that R_ξ , when developed in power series of ξ , of necessity only contains even powers of ξ , is not of universal application. With this latter end in view we reproduce (in abbreviated form) Taylor's development as it appeared in the important paper quoted above.

1. *Taylor's Deduction.*

Suppose we observe the values $p_1 \dots p_n$ of the pressure deviations as shown on a barogram at a large number of successive times $t_1 \dots t_n$. Suppose further that we observe $p_1 + \delta p_1 \dots p_n + \delta p_n$ at times $t_1 + \delta t \dots t_n + \delta t$, where δt is a small interval of time. Then, if $t_1 \dots t_n$ are taken at random, the mean value of p^2 is

$$[p^2] = (p_1^2 + p_2^2 + \dots + p_n^2)/n, \quad \dots \dots \dots (A)$$

and since we are considering a curve in which $[p^2]$ is constant, $[p^2]$ is also, to the first order, equal to

$$\frac{1}{n} \left\{ \left(p_1 + \frac{dp_1}{dt} \delta t \right)^2 \dots \left(p_n + \frac{dp_n}{dt} \delta t \right)^2 \right\} = [p^2] + 2 \left[p \frac{dp}{dt} \right] \delta t.$$

* Communicated by the Author.

† Taylor, G. I., "Diffusion by Continuous Movements," Proc. of the London Math. Soc. xx. p. 196 (1920).

Hence the condition that the standard deviation shall be constant leads to the result

$$\left[p \frac{dp}{dt} \right] = 0. \quad \dots \quad (B)$$

It appears, therefore, that we can differentiate within the square brackets, and if we differentiate (B), we find

$$\left[p \frac{d^2 p}{dt^2} \right] = - \left[\left(\frac{dp}{dt} \right)^2 \right] \dots \quad (C)$$

Similarly, it may be proved in general that

$$\left[p \frac{d^{2n} p}{dt^{2n}} \right] = (-1)^n \left[\left(\frac{d^n p}{dt^n} \right)^2 \right] \quad \text{and} \quad \left[p \frac{d^{2n+1} p}{dt^{2n+1}} \right] = 0. \quad \dots \quad (D)$$

If p be the value of p at time $t=0$, and p_ξ its value at $t=\xi$, then if p be measured as before at a large number of times $t_1 \dots t_n$, chosen at random, and p_ξ at times $t_1+\xi, \dots t_n+\xi$, then the correlation coefficient is by definition

$$R_\xi = [p p_\xi] / \{ [p^2] [p_\xi^2] \}^{\frac{1}{2}}.$$

But, by hypothesis, the standard deviation of p does not vary with the time, and hence

$$[p_\xi^2] = [p^2],$$

and

$$R_\xi = [p p_\xi] / [p^2]. \quad \dots \quad (E)$$

Now expand p_ξ in terms of ξ :

$$p_\xi = p + \xi \frac{dp}{dt} + \frac{\xi^2}{2!} \frac{d^2 p}{dt^2} + \dots$$

Hence
$$[p p_\xi] = [p^2] + \xi \left[p \frac{dp}{dt} \right] + \frac{\xi^2}{2!} \left[p \frac{d^2 p}{dt^2} \right] + \dots,$$

and finally, by making use of (D) and (E), we arrive at the result

$$R_\xi = 1 - \frac{\xi^2}{2! [p^2]} \left[\left(\frac{dp}{dt} \right)^2 \right] + \frac{\xi^4}{4! [p^2]} \left[\left(\frac{d^2 p}{dt^2} \right)^2 \right] - \dots \quad (F)$$

which means that R_ξ is an even function of ξ .

2. Criticism of Taylor's Deduction.

From actual measurements we know that, in general, the frequencies of atmospheric pressure and temperature deviations, after the regular daily and seasonal fluctuations have first been eliminated, arrange themselves according to the normal frequency curve, and the same frequency curve also applies to turbulent velocities in any given direction. Now the range of this distribution curve extends to infinity in both the

this is clearly equal to u times the function $f(p, \xi, u)$ integrated over all possible values of u :

$$[p_\xi] = \int_{-\infty}^{\infty} u f(p, \xi, u) du. \quad . \quad . \quad . \quad . \quad . \quad (2)$$

Since a given value of p can take on an infinite number of other values after a time ξ , instead of writing equation (E) as Taylor does, the meaning of R_ξ becomes much clearer if we write it as follows :

$$R_\xi = \frac{[p[p_\xi]]}{[p^2]}. \quad . \quad . \quad . \quad . \quad . \quad (3)$$

If now we consider not only the A molecules, but all the molecules in the gas, then it is clear that

$$[p[p_\xi]] = \int_{-\infty}^{\infty} p F(p) [p_\xi] dp,$$

and by virtue of (2) and (3)

$$R_\xi [p^2] = \int_{-\infty}^{\infty} \left\{ p F(p) \int_{-\infty}^{\infty} u f(p, \xi, u) du \right\} dp. \quad . \quad . \quad . \quad (4)$$

This is the general expression for the correlation coefficient, and its actual calculation thus depends upon our ability to determine the distribution function $f(p, \xi, u)$.

In the particular case where we have the normal distribution law as given in (1), and where $[p^2] = \alpha^2/2$, equation (4) reduces to

$$R_\xi = \frac{2}{\alpha^3 \sqrt{\pi}} \int_{-\infty}^{\infty} \left\{ p e^{-p^2/\alpha^2} \int_{-\infty}^{\infty} u f(p, \xi, u) du \right\} dp. \quad . \quad . \quad . \quad (5)$$

It need hardly be stressed that equation (5) is by no means confined to molecular velocities, but is applicable to such quantities as turbulent velocity, atmospheric temperature, pressure, etc., which follow the normal distribution law.

If we use the symbol $[[p_\xi^2]]$ in place of Taylor's $[p_\xi^2]$, reserving the latter to denote the mean value of the squares of the velocities of the A molecules only, then

$$[p_\xi^2] = \int_{-\infty}^{\infty} u^2 f(p, \xi, u) du. \quad . \quad . \quad . \quad . \quad . \quad (6)$$

Now since the mean value of the sum of the squares of the velocities of all the molecules remains constant and equal to $[p^2]$, it follows that

$$[[p_\xi^2]] = \int_{-\infty}^{\infty} \left\{ F(p) \int_{-\infty}^{\infty} u^2 f(p, \xi, u) du \right\} dp = [p^2]. \quad . \quad . \quad . \quad (7)$$

Having succeeded in determining the function $f(p, \xi, u)$, we can employ this last equation as a useful test for the correctness of our mathematical development.

4. The Derivation of the Fundamental Differential Equation.

An attempt was actually made to determine the value of R_ξ in the case of molecular velocities, and if certain simplifying assumptions are made it is found that R_ξ has the form $e^{-k\xi}$; but the mathematical difficulties proved to be very formidable indeed and it was not found possible to solve the problem satisfactorily. However, in the case of molecular velocities, the problem is completely specified, and the solution depends not upon any assumptions (except the assumption of perfect elastic impact between the molecules), but only upon our mathematical prowess.

On the other hand, although the readings of a barogram or thermogram in general obey the same distribution law as do molecular velocities, there is no known law, analogous to the law of elastic impact between molecules, which can serve as a basis for determining the value of R_ξ . Hence, in this case, the problem is not completely specified, and it thus becomes necessary to put forward some hypothesis on which to base our calculations.

In attempting to arrive at some reasonable hypothesis, let us consider the well-known case in the conduction of heat from a point source, which bears a striking resemblance to the problem under consideration. If there is a point source of heat of unit strength at $x=0$, and conduction takes place along a uniform infinite rod in the X direction, then the distribution of temperature T along the rod after a time t is given by the expression

$$T = \frac{1}{\beta\sqrt{\pi t}} e^{-x^2/\beta^2 t}, \quad \dots \dots \dots (8)$$

in which β is a constant depending upon the cross-section and thermal properties of the rod.

The differential equation from which (8) is derived is

$$\frac{dT}{dt} = k \frac{d^2T}{dx^2}, \quad \dots \dots \dots (9)$$

in which k is the thermal diffusivity of the material of the rod.

Equation (8) is of exactly the same form as (1), with this important difference, that the time t enters into the former in such a way that after a long lapse of time the heat becomes entirely dissipated along the rod, which means that the distribution curve becomes flat, whereas the distribution curve (1) is independent of the time. The question thus naturally arises whether entire similarity between the two cases cannot be attained by controlling the conduction of heat in such a way that the final temperature distribution curve will remain finite after infinite time.

This can indeed be attained by imagining an added "force" f , which at equilibrium exactly balances the conductive "force." Due to the

ordinary conduction of heat there is a transport of heat equal to $-K \frac{dT}{dx}$ across unit cross-section in unit time, K being the thermal conductivity of the rod. This transport proceeds as long as the temperature gradient $\frac{dT}{dx}$ persists. If, then, equilibrium is to be attained at finite values of $\frac{dT}{dx}$ at equilibrium, the same amount of heat must be transported in the opposite direction. If we denote by fT the heat transported by the extra "force" in unit time across unit cross-section, then

$$fT - K \frac{dT}{dx} = 0.$$

But, if T is to obey the normal distribution law at equilibrium,

$$T = \frac{1}{\alpha\sqrt{\pi}} e^{-x^2/\alpha^2},$$

it follows readily that

$$f = -\frac{2Kx}{\alpha^2} \dots \dots \dots (10)$$

The equation of the conduction of heat may now be deduced in the ordinary way by considering the total increase of heat in a small element dx in time dt , and equating this increase of heat to the rise in temperature multiplied by the appropriate constant. The final equation is found to be

$$\frac{\alpha^2}{k} \frac{dT}{dt} = \alpha^2 \frac{d^2T}{dx^2} + 2x \frac{dT}{dx} + 2T \dots \dots \dots (11)$$

Our hypothesis now consists in assuming that a continuously varying quantity, such as the barometric pressure at a point, atmospheric temperature, turbulent velocity etc., whose frequency distribution is normal, obeys a differential equation of the type (11). We thus adopt this differential equation as the basis for our further calculations.

It is freely conceded that this hypothesis cannot be defended by strictly logical reasoning, and the justification for its final adoption or rejection can only rest with the amount of agreement obtained between theory and experiment.

5. *The Solution of the Fundamental Equation.*

Adopting our former symbols and putting F for T and p for x in (11), our fundamental equation becomes

$$\frac{\alpha^2}{k} \frac{dF}{dt} = \alpha^2 \frac{d^2F}{dp^2} + 2p \frac{dF}{dp} + 2F \dots \dots \dots (12)$$

It can easily be verified that at equilibrium when $\frac{dF}{dt}=0$, F assumes the form given in (1).

In order to facilitate the solution of (12) we introduce the new variables x and ϕ , where

$$\left. \begin{aligned} p &= \alpha x, \\ F &= \phi e^{-x^2}, \end{aligned} \right\}, \quad \dots \dots \dots (13)$$

in which case (12) becomes

$$\frac{\alpha^2}{k} \frac{d\phi}{dt} = \frac{d^2\phi}{dx^2} - 2x \frac{d\phi}{dx} \dots \dots \dots (14)$$

If we now follow the usual procedure by assuming ϕ to be the product of a function T of t and a function y of x , i. e.,

$$\phi = yT, \quad \dots \dots \dots (15)$$

then substitution in (20) yields the two equations

$$\alpha^2 \frac{dT}{dt} = -2knT \dots \dots \dots (16)$$

and
$$\frac{d^2y}{dx^2} - 2x \frac{dy}{dx} + 2ny = 0, \quad \dots \dots \dots (17)$$

in which n is an arbitrary constant.

A solution of (16) is

$$T = e^{2knt/\alpha^2} \dots \dots \dots (18)$$

Provided n is zero or a whole number, we recognize in (17) the well-known Hermite equation, whose solution is

$$y = C_n H_n, \quad \dots \dots \dots (19)$$

where C_n is an arbitrary constant and the Hermite function H_n is defined as follows:

$$H_n = (-1)^n e^{x^2} \frac{d^n e^{-x^2}}{dx^n} \dots \dots \dots (20)$$

From the definition it follows that

$$\begin{aligned} H_0 &= 1, & H_1 &= 2x, \\ H_2 &= 2 - 4x^2, & H_3 &= 12x - 8x^3, \text{ etc.} \end{aligned}$$

Making use of the fact that the sum of any number of solutions is also a solution, and remembering that $F = Ty e^{-x^2}$ and $p = \alpha x$,

$$F = e^{-x^2/\alpha^2} \sum_{n=0}^{\infty} C_n H_n \left(\frac{p}{\alpha} \right) e^{-2knt/\alpha^2} \dots \dots \dots (21)$$

This is a solution of our fundamental equation (12), and if the infinite number of constants C_n can be chosen so as to satisfy any given boundary conditions, then we have arrived at the solution of our problem.

Suppose that

$$F = F_0(p) \text{ when } t = 0. \quad (22)$$

Due to the orthogonal properties of the H functions it is possible to deduce the values of C_n ,

$$\text{for } \int_{-\infty}^{\infty} e^{-x^2} H_n(x) H_m(x) dx = 0 \text{ when } m \neq n \\ = 2^n n! \sqrt{\pi} \text{ when } m = n,$$

from which it follows that

$$C_n = \frac{1}{2^n n! \alpha \sqrt{\pi}} \int_{-\infty}^{\infty} F(p) H_n\left(\frac{p}{\alpha}\right) dp. \quad (23)$$

We have now arrived at the stage where it is possible for us to determine the function $f(p, \xi, u)$. If we call to mind that $f(p, \xi, u)$ is the distribution function of the A molecules whose initial velocities lay between p and $p + dp$, which amounts to the same as the distribution of a unit source initially situated at p , it becomes clear that we are dealing with the case where

$$F_0(p) = 0 \text{ from } -\infty \text{ to } p, \\ F_0(p) dp = 1 \text{ from } p \text{ to } p + dp, \\ F_0(p) = 0 \text{ from } p + dp \text{ to } \infty,$$

and hence equation (23) yields the result

$$C_n = H_n\left(\frac{p}{\alpha}\right) / 2^n n! \alpha \sqrt{\pi}.$$

Remembering that we employ the symbol u instead of p when dealing with the selected A molecules, we find that equation (21) becomes

$$f(p, \xi, u) = \frac{e^{-u^2/\alpha^2}}{\alpha \sqrt{\pi}} \sum_{n=0}^{\infty} \frac{H_n\left(\frac{p}{\alpha}\right) H_n\left(\frac{u}{\alpha}\right)}{2^n n!} e^{-2kn\xi/\alpha^2}. \quad (24)$$

In fig. 1 this function is shown in graphical form as computed from the above equation. The point source was originally situated at $\frac{p}{\alpha} = 1$.

The curve on the right shows the distribution of this point source after a time $\xi = 0.347 \alpha^2/k$. The curve next to it gives the distribution when $\xi = 0.549 \alpha^2/k$, and the third curve is the normal distribution curve which is attained after infinite time. It will be seen that with the passage of time the distribution curve becomes less and less steep, at the same

time moving to the left until it finally assumes the normal symmetrical form about the co-ordinate axis.

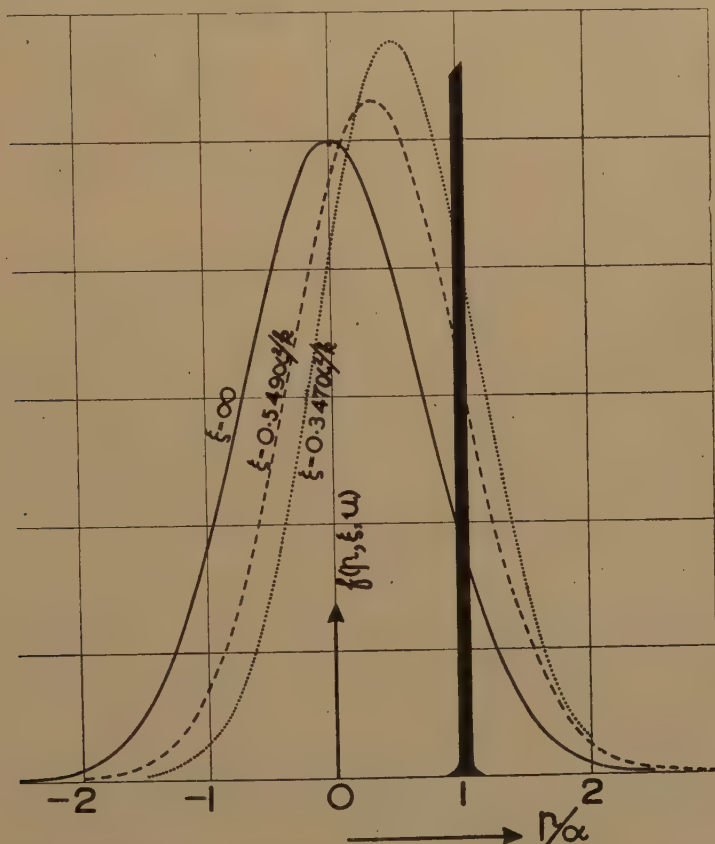
If the above value of $f(p, \xi, u)$ be substituted in (2) it is found that

$$[p_\xi] = pe^{-2k\xi/a^2}, \quad \dots \dots \dots (25)$$

a consequence of the fact that

$$\int_{-\infty}^{\infty} xe^{-x^2} H_n(x) dx = \sqrt{\pi}, \text{ when } n=0 \text{ for all other values of } n.$$

Fig. 1.



The distribution curves of a point source situated at $p/a=1$ (represented by the heavy vertical line) after intervals of time $\xi=0.347a^2/k$, $\xi=0.549a^2/k$, $\xi=\infty$.

It thus follows from (5) that

$$R_\xi = \frac{2}{\alpha^3 \sqrt{\pi}} \int_{-\infty}^{\infty} p \exp. \left(-\frac{p^2 + 2k\xi}{\alpha^2} \right) dp = e^{-2k\xi/a^2}. \quad \dots \dots (26)$$

The correlation coefficient has thus been found to be a simple exponential function of the time, which does not conform to Taylor's demand that its expansion should contain only even powers of ξ .

Finally, it is of interest to note that equation (7) can be directly verified ; for

$$[p]_{\xi}^2 = \alpha^2/2 + (p^2 - \alpha^2/2)e^{-4k\xi/\alpha^2},$$

which follows from the relation

$$\begin{aligned} \int_{-\infty}^{\infty} x^2 e^{-x^2} H_n(x) dx &= 2\sqrt{\pi} \text{ when } n=2, \\ &= \sqrt{\pi}/2 \text{ when } n=0, \\ &= 0 \text{ for all other values of } n. \end{aligned}$$

Hence, according to equation (7),

$$[[p_{\xi}^2]] = \frac{1}{\alpha\sqrt{\pi}} \int_{-\infty}^{\infty} \left\{ \alpha^2/2 + \left(p^2 - \frac{\alpha^2}{2} \right) \exp. (-4k\xi/\alpha^2) \right\} e^{-p^2/\alpha^2} dp = \alpha^2/2 = [p^2],$$

as should be the case.

6. *Experimental Results.*

Although the theory developed above is self-sufficient as a theory, and adequately serves to illustrate the serious limitations of Taylor's deduction, yet considerable interest must attach to the comparison of the theory with experiment.

The calculation of correlation coefficients to a sufficient degree of constancy entails a considerable amount of work. However, a few coefficients were actually computed for a couple of South African stations, and it is hoped to extend this work in the near future.

First Experiment.—Daily minimum temperatures at Kimberley (lat. $28^{\circ} 44'$ S., long. $24^{\circ} 46'$ E.) for the period 1931 to 1935 inclusive.

In Table I. the values of R_{ξ} , where $\xi=1, 2, 3, 4$, and 5 days, are given, together with the corresponding values of the theoretical curve which fit the data best.

TABLE I.—Showing Experimental and Theoretical Values of the Correlation Coefficient.

No. of days.	January–June.		July–December.	
	R (exp.).	R (theor.).	R (exp.).	R (theor.)
1	·643	·638	·557	·537
2	·393	·408	·239	·288
3	·225	·261	·175	·155
4	·173	·167	·093	·083
5	·150	·107	·045	·045

Second Experiment.—The corresponding figures for Groot Drakenstein (lat. $33^{\circ} 51' S.$, long. $18^{\circ} 57' E.$) computed for the same period are contained in Table II.

The results show quite fair agreement between theory and experiment, and there can be little doubt that if a longer series of data had been employed the correspondence would have been closer. For the computation of each correlation coefficient a series of 913 temperatures was used, but it seems advisable to use about ten times this number in order to reduce the probable error to reasonable proportions.

However, the theory cannot be applied, without modification, to any time series whose standard deviation is not constant. If, *e. g.*, we subtract the hourly temperature readings at any station from the mean hourly temperatures, we obtain a series containing hourly temperature

TABLE II.

January–June.			July–December.	
No. of days.	R (exp.).	R (theor.).	R (exp.).	R (theor.).
1	.577	.602	.563	.520
2	.372	.362	.249	.270
3	.229	.218	.119	.140

departures, but in general the dispersion of these departures is much greater in the afternoon than in the early hours of the morning. In other words, the standard deviation varies sharply with the time of day. Hence, if we attempt to calculate correlation coefficients for periods of less than 24 hours, we are dealing with a series where the standard deviation, α , is by no means constant, and obviously the theory developed above applies only to a *homogeneous* series in which α is constant.

The auto-correlation coefficient, as it is generally called, was calculated by K. Stumpff* for hourly pressure readings at Potsdam, and he also found that it approximately decreases exponentially with the time. In the case of pressure readings, however, there is generally one or more harmonic components of appreciable amplitude, and it is therefore not surprising that Stumpff's results show a slightly negative value for R

* K. Stumpff, "Ueber die Zufallswahrscheinlichkeit von Periodizitaeten," Veroff. Met. Inst. Univ. Berlin, no. 4, 1936.

after the fifth day. In order to obtain a "homogeneous" series of pressure readings, any possible harmonic component must first be removed.

At present a fairly comprehensive study is being carried out in connexion with the auto-correlation of temperature and pressure series, and the results will be published elsewhere in due course.

My thanks are due to Prof. J. T. Morrison for his kindly criticism of this paper.

XLVI. *On the Solution of Second Order Differential Equations satisfying Boundary Conditions.*

By W. R. MORGANS, M.Sc.*

[Received September 3, 1941.]

THE following note has arisen from the consideration of the distribution of sound in front of a spherical concave reflecting surface.

In wave mechanics and in many branches of mathematical physics a function V is required to satisfy an equation of the type

$$\nabla^2 V + [\lambda + f(x, y, z)]V = 0.$$

From the point of view of electrostatics†, the term $[\lambda + f(x, y, z)]V$ may be considered as a density of electric charge (from Poisson's equation), and the question arises whether the boundary problem can be solved by replacing this volume density by a series of charged surfaces and proceeding to the limit.

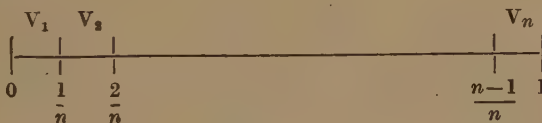
With this physical analogy, the following note shows that the correct eigen values are obtained by this limit process for the following particular cases of the above equation.

i. $\frac{d^2 V}{dx^2} + \kappa^2 V = 0.$ (a) $V = 0$ when $x = 0, 1.$
(b) $V = 1$ when $x = 0, 1.$

ii. $\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \kappa^2 V = 0.$ $V = 0$ on the boundary of a square.

iii. $\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \frac{\partial^2 V}{\partial z^2} + \kappa^2 V = 0.$ $V = 0$ on the boundary of a cube.

i. (a). The equation $\frac{d^2 V}{dx^2} + \kappa^2 V = 0$ with the boundary conditions $V = 0$ when $x = 0$ and $1.$



* Communicated by the Author.

† Cf. Courant and Hilbert, 'Methoden der Mathematischen Physik,' i. pp. 273 et seq.

Introduce charged surfaces at

$$x = \frac{1}{n}; \frac{2}{n}; \dots; \frac{n-1}{n},$$

and the boundary conditions

$$\left. \begin{aligned} V_r &= V_{r+1}, \\ n \left(\frac{dV_{r+1}}{dx} - \frac{dV_r}{dx} \right) + \kappa^2 V_r &= 0, \end{aligned} \right\} \text{at } x = \frac{r}{n}.$$

We seek solutions V_r , such that $\frac{d^2 V_r}{dx^2} = 0$ everywhere, and such that the preceding boundary conditions are satisfied.

$$\begin{aligned} \text{Let } V_r &= xA_r + B_r \quad r > 1, \\ V_1 &= x. \end{aligned}$$

It can be shown that

$$\begin{aligned} A_{r+1} &= A_r - \frac{\kappa^2}{n^2} [A_1 + A_2 + \dots + A_r], \\ B_{r+1} &= -\frac{r}{n} A_{r+1} + \frac{1}{n} [A_1 + A_2 + \dots + A_r]. \end{aligned}$$

$$\text{Let } f(t) = A_1 + A_2 t + A_3 t^2 + \dots,$$

then it can be shown that

$$f(t) = \frac{1-t}{1-2t \cos \theta + t^2}$$

$$\text{if } 1 - \frac{\kappa^2}{2n^2} = \cos \theta.$$

It can be shown that

$$V_{r+1} = \left(x - \frac{r}{n} \right) \left[\cos r\theta + \tan \frac{\theta}{2} \sin r\theta \right] + \frac{\sin r\theta}{n \sin \theta}.$$

The condition that $V_{r+1} = 0$ when $r = n$ is satisfied if

$$\sin n\theta = 0.$$

Hence the possible values of κ are given by

$$1 - \frac{\kappa^2}{2n^2} = \cos \frac{s\pi}{n},$$

where s is an integer. Thus

$$\kappa = 2n \sin \frac{s\pi}{2n}. \quad \dots \dots \dots (1)$$

When $n \rightarrow \infty$, equation (1) gives

$$\kappa = s\pi.$$

It is seen, therefore, that in this case the limit process is successful as it gives the correct eigen values.

i. (b). The equation $\frac{d^2V}{dx^2} + \kappa^2 V = 0$ with the boundary conditions $V=1$ when $x=0$ and 1.

A similar method of solution can be applied in this case with the assumption that

$$V_r = xA_r + B_r, \quad r > 1,$$

$$V_1 = 1.$$

Here it can be shown that

$$f(t) = \frac{-2n(1 - \cos \theta)t}{1 - 2t \cos \theta + t^2}$$

and
$$V_r = 2n \left(\frac{r}{n} - x \right) \tan \frac{\theta}{2} \sin (r-1)\theta + \frac{\cos \left(\frac{2r-1}{2} \theta \right)}{\cos \frac{\theta}{2}},$$

where θ is defined as in case i. (a).

The condition that $V_r = 1$ when $r = n$ is satisfied if

$$\theta = \frac{2s\pi}{n} \quad \text{or} \quad \frac{2s\pi}{n-1},$$

where s is an integer. Therefore, the possible values of κ are given by

$$\kappa = 2n \sin \frac{s\pi}{n} \quad \text{or} \quad 2n \sin \frac{s\pi}{n-1}.$$

When $n \rightarrow \infty$, both of these give $\kappa = 2s\pi$, which are the correct eigen values.

ii. The partial differential equation

$$\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \kappa^2 V = 0,$$

with the boundary conditions

$$V=0, \quad x=0 \text{ and } x=1 \text{ for all } y,$$

$$V=0, \quad y=0 \text{ and } y=1 \text{ for all } x.$$

A similar method of solution can be adopted as in i. (a), with the charged surfaces introduced along

$$x = \frac{1}{n}, \frac{2}{n}, \dots, \frac{n-1}{n},$$

$$y = \frac{1}{m}, \frac{2}{m}, \dots, \frac{m-1}{m},$$

and the boundary conditions

$$\left. \begin{aligned} &V_{r,p} = V_{r+1,p}, \\ &n \left(\frac{\partial V_{r+1,p}}{\partial x} - \frac{\partial V_{r,p}}{\partial x} \right) + \lambda^2 V_{r,p} = 0, \end{aligned} \right\} \text{at } x = \frac{r}{n} \text{ for all } y,$$

$$\left. \begin{aligned} &V_{r,p} = V_{r,p+1}, \\ &m \left(\frac{\partial V_{r,p+1}}{\partial y} - \frac{\partial V_{r,p}}{\partial y} \right) + \mu^2 V_{r,p} = 0, \end{aligned} \right\} \text{at } y = \frac{p}{m} \text{ for all } x,$$

where

$$\lambda^2 + \mu^2 = \kappa^2; \quad \lambda, \mu \neq 0.$$

We seek solutions $V_{r,p}$ such that $\nabla^2 V_{r,p} = 0$ everywhere. Let

$$\begin{aligned} V_{r,p} &= A_{r,p} + B_{r,p}x + C_{r,p}y + D_{r,p}xy, & r, p > 1, \\ V_{1,p} &= B_{1,p}x + D_{1,p}xy, & r = 1, p > 1, \\ V_{r,1} &= C_{r,1}y + D_{r,1}xy, & r > 1, p = 1, \\ V_{1,1} &= xy, & r, p = 1. \end{aligned}$$

Then, as in case i. (a), it is not difficult to show that

$$V_{r,p} = \left[\left(x - \frac{r}{n} \right) \left(\cos r\theta + \tan \frac{\theta}{2} \sin r\theta \right) + \frac{\sin r\theta}{n \sin \theta} \right] \left[\left(y - \frac{p}{m} \right) \right. \\ \left. \times \left(\cos p\phi + \tan \frac{\phi}{2} \sin p\phi \right) + \frac{\sin p\phi}{m \sin \phi} \right],$$

where

$$1 - \frac{\lambda^2}{2n^2} = \cos \theta; \quad 1 - \frac{\mu^2}{2m^2} = \cos \phi.$$

The condition $V_{r,p} = 0$ when $r = n, x = 1$ is satisfied if $\sin n\theta = 0$. Hence the possible values of λ are given by

$$\lambda = 2n \sin \frac{s\pi}{2n}, \quad \lambda \neq 0, \quad s = \pm 1, \pm 2, \text{ etc.}$$

Similarly, the possible values of μ are given by

$$\mu = 2m \sin \frac{t\pi}{2m}, \quad \mu \neq 0, \quad t = \pm 1, \pm 2, \text{ etc.}$$

When $n \rightarrow \infty$, or $m \rightarrow \infty$, these equations give

$$\lambda = s\pi, \quad s = \pm 1, \pm 2, \text{ etc.}$$

$$\mu = t\pi, \quad t = \pm 1, \pm 2, \text{ etc.}$$

Hence the possible values of κ are given by

$$\kappa^2 = \lambda^2 + \mu^2 = (s^2 + t^2)\pi^2,$$

where s and t are integers different from zero. The limit process is again successful.

iii. The equation

$$\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \frac{\partial^2 V}{\partial z^2} + \kappa^2 V = 0$$

with $V=0$ on the boundary of a unit cube.

The function $V_{r,p,q}$ required for the three-dimensional problem can best be obtained by consideration of the function $V_{r,p}$ of the two-dimensional problem. This function $V_{r,p}$ can be obtained simply from the solution V_{r+1} of case i. (a). If we consider the function $V_{r,p}$ defined by the equations.

$$V_{r,p} = V_r V_p,$$

where

$$V_r = \left(x - \frac{r}{n}\right) \left(\cos r\theta + \tan \frac{\theta}{2} \sin r\theta\right) + \frac{\sin r\theta}{n \sin \theta},$$

$$V_p = \left(y - \frac{p}{m}\right) \left(\cos p\phi + \tan \frac{\phi}{2} \sin p\phi\right) + \frac{\sin p\phi}{m \sin \phi},$$

then it is easy to show that $V_{r,p}$ thus defined satisfies all the conditions of problem ii., provided $\lambda^2 + \mu^2 = \kappa^2$; $\lambda, \mu \neq 0$.

Similarly, for the three-dimensional problem, we obtain the function

$$V_{r,p,q} = \left\{ \left(x - \frac{r}{n}\right) \left(\cos r\theta + \tan \frac{\theta}{2} \sin r\theta\right) + \frac{\sin r\theta}{n \sin \theta} \right\} \\ \times \left\{ \left(y - \frac{p}{m}\right) \left(\cos p\phi + \tan \frac{\phi}{2} \sin p\phi\right) + \frac{\sin p\phi}{m \sin \phi} \right\} \\ \times \left\{ \left(z - \frac{q}{l}\right) \left(\cos q\psi + \tan \frac{\psi}{2} \sin q\psi\right) + \frac{\sin q\psi}{l \sin \psi} \right\},$$

where $\kappa^2 = \lambda^2 + \mu^2 + \nu^2$,

and where the possible values of λ, μ, ν are given by

$$1 - \frac{\lambda^2}{2n^2} = \cos \frac{s\pi}{n}; \quad 1 - \frac{\mu^2}{2m^2} = \cos \frac{t\pi}{m}; \quad 1 - \frac{\nu^2}{2l^2} = \cos \frac{w\pi}{l},$$

s, t, w being integers different from zero.

If $n, m, l \rightarrow \infty$, the possible values of λ, μ, ν are given by

$$\lambda = s\pi, \quad \mu = t\pi, \quad \nu = w\pi,$$

where

$$\kappa^2 = (s^2 + t^2 + w^2)\pi^2.$$

The function $V_{r,p,q}$ can also be obtained from first principles.

iv. It may be noted that the charged surfaces need not be equidistant, but it happened in the cases chosen that the choice was the most suitable. For example, in two dimensions, we may consider the space divided into a network by orthogonal surfaces $\xi = \xi_r, \eta = \eta_p$ defined by the orthogonal transformation $x = x(\xi, \eta), y = y(\xi, \eta)$. The differential equation

$$\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \kappa^2 V = 0$$

is transformed into

$$h_1 h_2 \left[\frac{\partial}{\partial \xi} \left(\frac{h_1}{h_2} \frac{\partial V}{\partial \xi} \right) + \frac{\partial}{\partial \eta} \left(\frac{h_2}{h_1} \frac{\partial V}{\partial \eta} \right) \right] + \kappa^2 V = 0,$$

where the line element is given by

$$ds^2 = \frac{d\xi^2}{h_1^2} + \frac{d\eta^2}{h_2^2}.$$

We seek solutions $U_q(\xi, \eta)$ of the equation

$$\frac{\partial}{\partial \xi} \left(\frac{h_1}{h_2} \frac{\partial U}{\partial \xi} \right) + \frac{\partial}{\partial \eta} \left(\frac{h_2}{h_1} \frac{\partial U}{\partial \eta} \right) = 0,$$

such that the function

$$V_{r,p}(\xi, \eta) = A_{r,p} + \sum_a B_{r,p}^a U_a(\xi, \eta)$$

satisfies the conditions

$$\left. \begin{aligned} V_{r,p} &= V_{r+1,p}, \\ \frac{1}{p_{\xi r}} \left[\frac{\partial V_{r+1,p}}{\partial s_\xi} - \frac{\partial V_{r,p}}{\partial s_\xi} \right] + \lambda^2 V_{r,p} &= 0, \end{aligned} \right\} \text{ at } \xi = \xi \text{ all } \eta, \text{ etc..}$$

where

$$\begin{aligned} \frac{\partial}{\partial s_\xi} &= h_1 \frac{\partial}{\partial \xi}, & \frac{\partial}{\partial s_\eta} &= h_2 \frac{\partial}{\partial \eta}, \\ p_{\xi r} &= \int_{\xi_r}^{\xi_{r+1}} ds_\xi; & p_{\eta p} &= \int_{\eta_p}^{\eta_{p+1}} ds_\eta, \\ ds_\xi &= \frac{d\xi}{h_1}, & ds_\eta &= \frac{d\eta}{h_2}. \end{aligned}$$

When the function $V_{r,p}$ is also chosen to satisfy the boundary conditions along ξ or η surfaces, the eigen values are then determined by a limit process.

The method appears to give a quick method of approximating to the eigen values of differential equations.

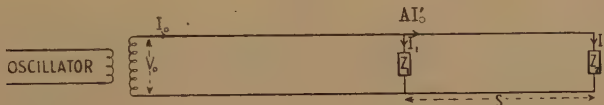
XLVII. A Method for the Measurement of Impedance by means of Lecher Wires.

By H. T. FLINT and GORDON WILLIAMS, King's College, London, at the H. H. Wills Physics Laboratory, The University, Bristol *.

[Received July 26, 1941.]

IN the course of investigations into a method for determining the conductivity and dielectric constant of solutions by means of Lecher wires, it became necessary to eliminate variations in the input electromotive force responsible for the stationary waves set up in the system. This led to the examination of a method in which the ratios of the currents or voltages at two points of the parallel wires were measured, these ratios being independent of any fluctuations of input provided that they

Fig. 1.



do not occur during the short space of time required to read the meters inserted into the circuit.

In the end it was found convenient to measure the currents flowing in a bridge placed across the distant end of the wires and in a bridge placed between this end and the input end. The distance between the two bridges is variable, and the fact that a simple relation exists between the ratio of the currents in the bridges and their distance apart makes the method a practical one. It would appear that the method could be adapted for the original purpose of investigating properties of solutions, but it has been applied for the purpose of measuring the impedance of a current meter, for which it appears to be particularly suitable.

The Theory of the Method.

Referring to fig. 1 for the explanation of the notation, the impedance of the circuit beyond A is

$$Z_0 \cdot \frac{Z_2 \cosh Ps + Z_0 \sinh Ps^{(1)}}{Z_0 \cosh Ps + Z_2 \sinh Ps} \dots \dots \dots (1)$$

* Communicated by the Authors.

Thus
$$\frac{I_1}{I_0'} = \frac{Z_0}{Z_1} \cdot \frac{Z_2 \cosh Ps + Z_0 \sinh Ps}{Z_0 \cosh Ps + Z_2 \sinh Ps} \quad (2)$$

In these expressions P denotes the propagation constant and, in the present case, the resistance and leakance of the wires is neglected.

Thus P is an imaginary quantity $i\beta$, where $\beta = \omega\sqrt{LC} = 2\pi/\lambda$, L denoting the inductance and C the capacity per unit length of the line, while $w = 2\pi \times$ frequency and λ denotes the wave-length of the electromagnetic waves. The quantity Z_0 denotes the characteristic impedance of the wires and, in this case,

$$Z_0 = \sqrt{L/C}. \quad (3)$$

For a pair of parallel wires of diameter d , with an axis separation denoted by D ,

$$Z_0 = 4 \log_e 2D/d, \quad (4)$$

the medium round the wires being air.

Under these circumstances the hyperbolic functions become simple trigonometrical functions.

The current at any point in the length s at a distance x from A is given by the formula

$$I = I_0' \frac{Z_2 \sinh P(s-x) + Z_0 \cosh P(s-x)^{(2)}}{Z_2 \sinh Ps + Z_0 \cosh Ps} \quad (5)$$

Thus

$$I_2 = \frac{Z_0 I_0'}{Z_2 \sinh Ps + Z_0 \cosh Ps} = \frac{Z_1 I_1}{Z_2 \cosh Ps + Z_0 \sinh Ps}, \quad (6)$$

$$\frac{I_1}{I_2} = \frac{1}{Z_1} (Z_2 \cosh Ps + Z_0 \sinh Ps). \quad (7)$$

It is convenient to introduce the reflexion coefficient $K_2 = (Z_0 - Z_2)/(Z_0 + Z_2)$. Thus the expression in the bracket of equation (7) becomes

$$Z_0(e^{Ps} - K_2 e^{-Ps})/(1 + K_2).$$

Writing $K_2 = e^{-2r}$, this expression takes the form

$$Z_0 e^{-r} \{e^{(Ps+r)} - e^{-(Ps+r)}\}.$$

From these results it follows that

$$I_1/I_2 = 2Z_0 e^{-r} \sinh (Ps+r)/Z_1(1+K_2). \quad (8)$$

In the measurement of this current ratio it is the modulus, ρ , of the ratio which is determined. For different values of s its variation is determined by the hyperbolic function, all other quantities being constant in equation (8).

The quantity r is in general complex and is denoted by $(a+ib)$.

Thus

$$Ps+r=a+i(b+\beta s)$$

$$\begin{aligned} |\sinh (Ps+r)|^2 &= \sinh^2 a \cos^2(b+\beta s) + \cosh^2 a \sin^2(b+\beta s) \\ &= \sinh^2 a + \sin^2(b+\beta s), \end{aligned}$$

and ρ^2 is of the form

$$\rho^2 = A(\sinh^2 a + \sin^2(b+\beta s)), \quad (9)$$

A denoting a constant.

The maxima of ρ occur for values of s such that $(b+\beta s) = \frac{\pi}{2}, \frac{3\pi}{2}$, etc.

If the value of ρ , for all these cases, is denoted by ρ_m ,

$$\rho_m^2 = A \cosh^2 a. \quad (10)$$

From this it follows that

$$1 - \left(\frac{\rho}{\rho_m} \right)^2 = \frac{\cos^2(b+\beta s)}{\cosh^2 a}. \quad (11)$$

The minima of ρ occur when $(b+\beta s) = 0, \pi, 2\pi$, etc., and the value of ρ^2 at these points is $A \sinh^2 a$ or, by equation (10),

$$\rho_{\min} = \rho_m \tanh a. \quad (12)$$

Equation (11) shows that a graph of ρ drawn against s is symmetrical about the turning-points. This is an advantage in the determination of the exact values of s at these points.

The distance between successive maxima and minima is $\lambda/4$, so that the determination of λ can be made accurately by means of the graph. It is to be noted that the ratio ρ/ρ_m occurs in this relation and, since ρ denotes the ratio of the currents measured by the meters, this fact makes it unnecessary to know the meter constant. All that is required is the law of the relation between the current and deflexion. In an experiment to be described, the square of the current was proportional to the deflexion.

In order to make use of the formulæ ρ is plotted against s , the curve obtained being illustrated in fig. 2.

The exact positions of the turning-points P and Q can be found by drawing horizontal lines AB and CD and, relying upon the symmetry of the curve, by measuring the distance between the mid-points of AB and CD. This distance is $\lambda/4$, so that λ can be determined.

From the fact that the first maximum occurs when

$$b+\beta s = \pi/2,$$

and that s is determined from the mid-point of CD, it is possible to obtain b . The value of this constant may also be determined from the position of the first minimum.

Equation (11) shows that, if ρ^2 be plotted against $\cos^2(b+\beta s)$, a linear graph results, from which the constant a may be determined. This graph is illustrated in fig. 3; the intercept OA gives the value of ρ_m^2 and the intercept OB that of $\cosh^2 a$. Thus the quantities a and b are determined and the impedance Z_2 can be expressed in terms of them and of Z_0 .

Fig. 2.

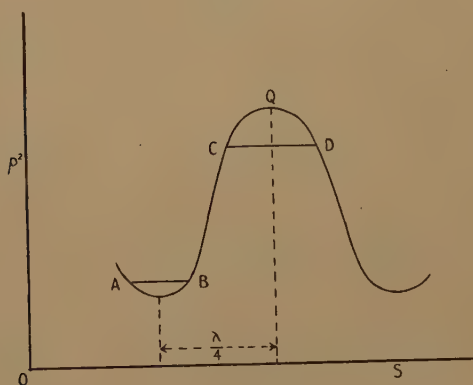
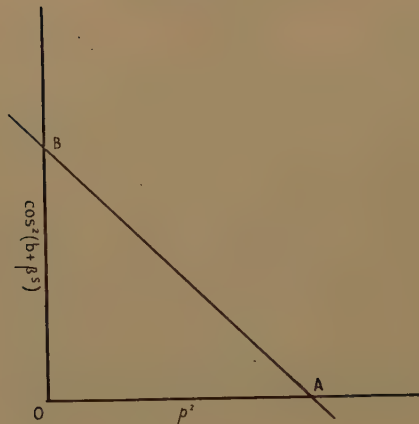


Fig. 3.



Expressions for the Resistive and Reactive Components of Z_2 :

In terms of the notation already introduced

$$K_2 = (Z_0 - Z_2)/(Z_0 + Z_2) = e^{-2r} = e^{-2(a+ib)},$$

so that

$$Z_2 = Z_0 \tanh(a+ib).$$

If Z_0 is purely resistive, as is supposed in the present case, the real or resistive component of Z_2 is

$$Z_0 \tanh a \sec^2 b / (1 + \tanh^2 a \tan^2 b),$$

and the imaginary or reactive component is

$$Z_0 \operatorname{sech}^2 a \tan b / (1 + \tanh^2 a \tan^2 b).$$

The Application of these Results in the Determination of an Impedance.

The method has been applied in the determination of the impedance of a thermal detector such as is used in the measurement of currents of radio frequency.

The waves were generated by means of a valve oscillator of the Hartley type. The wave-length was of the order of 200 cm., and in the case to be described it was 224 cm.

Attenuation in the wires was neglected for the purpose of this investigation, but it can be considered in the same way as in Chipman's experiments⁽³⁾.

The apparatus was used without any special precautions with regard to shielding, but the observations were made in a laboratory in which the observer maintained the same position when taking the readings from the meters.

Even under these conditions it appears that with this wave-length the method is capable of giving results of a high degree of accuracy. The results given here are of one of a series of observations made with different meters, which happened to be available in the laboratory.

The arrangement of apparatus is illustrated in fig. 1. The impedances Z_1 and Z_2 in this case are those of thermal detectors, and they are connected to microammeters possessing a square law scale. Thus the readings readily give ρ^2 and their ratios are plotted since these are equal to this quantity.

The readings in the experiment were taken for 5 cm. displacements of Z_1 between the values for s measured from Z_2 ranging from 30 cm. to 130 cm.

The table of values on p. 494 was obtained, and from them the symmetrical curve of ρ^2 against s was drawn (fig. 4).

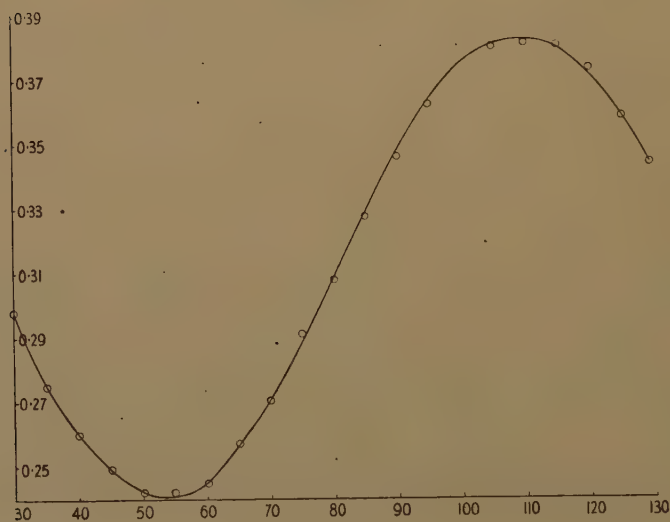
In fig. 5 values of ρ^2 have been plotted against the square of the cosine term. In this case the scale is rather large and the deviations from the line are not greater than those which would result from inaccuracies introduced in reading the microammeters.

The least reliable of the points on the linear graph are those corresponding to values of the ratio near the maximum and minimum which lie at the extremities of the line. But between them the values can be obtained repeatedly with great reliability. The curve of fig. 4 gives the value 224 cm. for the wave-length. The value of s at the maximum is

109.5 cm. and the value of b calculated from $(b+\beta s)=\pi/2$ is consequently -86° . The last two columns of the table can be completed from these values of λ and b .

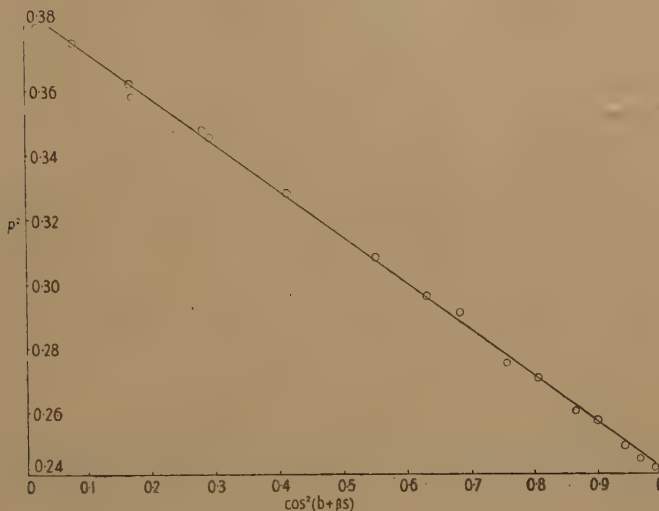
s cm.	I_1^2	I_2^2	ρ^2	$2\pi s/\lambda+b$	$\cos^2 (2\pi s/\lambda+b)$
30	23.4	79.0	.296	-37.8	.624
35	22.7	82.5	.275	-29.8	.753
40	22.6	87.0	.260	-21.8	.863
45	22.4	90.0	.249	-13.8	.943
50	23.0	95.0	.242	- 5.8	.990
55	23.3	96.5	.242	2.3	.998
60	24.2	98.5	.245	10.5	.966
65	25.3	98.5	.257	18.5	.899
70	26.2	97.0	.270	26.5	.801
75	27.2	93.5	.291	34.4	.681
80	27.5	89.0	.308	42.3	.548
85	27.8	84.5	.328	50.3	.408
90	27.6	79.5	.347	58.3	.277
95	27.2	75.0	.362	66.2	.163
100	26.6	71.0	.375	74.2	.074
105	26.0	68.5	.380	82.2	.018
110	25.5	67.0	.381	90.2	.000
115	25.1	66.0	.380	98.2	.020
120	24.6	66.0	.373	106.2	.078
125	24.0	67.0	.358	114.2	.168

Fig. 4.



From the linear graph the intercept on the horizontal axis is found to be 2.68. This is the value of $\cosh^2 a$.

Fig. 5.



There is thus a small reactive component corresponding to $b = -86^\circ$ and it is capacitive. The chief component of the reactance is resistive and the numerical values obtained lead to the value 414 ohms for this component. The calculation requires a knowledge of the characteristic resistance Z_0 , which, in this case, was 328 ohms.

References.

- (1) J. A. Fleming, 'The Propagation of Electric Currents in Telephone and Telegraph Conductors,' 4th edn., p. 99, equation 61.
- (2) *Loc. cit.* pp. 98, 99, equations 55 and 59.
- (3) Journ. App. Physics. x. p. 27 (1939); L. Hartshorn, 'Radio-Frequency Measurements, p. 243 *seqq.* (1940).

XLVIII. *The Shot Effect in a Saturated Diode.*

By N. R. CAMPBELL, Sc.D., F.Inst.P., and
V. J. FRANCIS, B.Sc., A.R.C.S., F.Inst.P., A.M.I.E.E.*

(Communication from the Research Staff of the M.O. Valve Co., Ltd.,
at the G.E.C. Research Laboratories, England.)

[Received August 5, 1941.]

Summary.

IN a previous paper (Phil. Mag. xxxii. p. 239 (1941)) we proved that Campbell's theorem and Schottky's theorem lead to the same result in calculating the shot noise in a valve, so long as (a) the valve is a saturated diode, (b) the transit time of the electrons is infinitely small, (c) the circuit impedances are all lumped. In this paper restrictions (b) and (c) are removed; restriction (a), though probably unnecessary, cannot be removed until the theory of the "space-charge effect" on noise is completely understood.

Since the methods adopted in this paper are rather different from those used previously, it has been thought better to make the papers formally independent and not to refer to the earlier paper for the small amount of matter common to both.

1. There are two alternative methods of calculating shot noise. One is based on Campbell's theorem †

$$\overline{\theta^2} - (\overline{\theta})^2 = N\epsilon^2 \int_0^\infty f^2(t) dt. \quad \dots \dots \dots (1.1)$$

The other is Schottky's theorem †

$$\overline{i^2} = 2I\epsilon\delta f, \quad \dots \dots \dots (1.2)$$

where ϵ is the electronic charge.

The purpose of this note is to show that, when the shot noise is characteristic of a saturated diode, the two methods are equivalent and must always lead to the same result. The treatment of any discharge device other than a saturated diode requires the introduction of additional

* Communicated by C. C. Paterson.

† N. R. Campbell, Proc. Camb. Phil. Soc. xv. pp. 117–136, 310–328 (1909). We have not ascertained who first stated the theorem precisely in this form, but it is always regarded as based on W. Schottky, *Ann. d. Phys.* lvii. p. 541 (1918).

physical hypotheses. There appears to be no doubt that, if the same hypothesis is associated with each theorem, the result will again be the same ; but we shall attempt no formal proof of this proposition.

2. Neither theorem is completely general or a complete expression of the conceptions underlying it. Our first task then will be to generalize each proposition ; then only will it be possible to put both of them in identical mathematical form to show that the symbols in each are physically identical, and thus to prove the equivalence.

3. We start with Campbell's theorem. The basic conception here is that electrons are passing instantaneously between the electrodes at a constant mean rate N , but otherwise at random instants. θ is the reading at time t of some instrument connected to the electrodes, whose nature is such that, if a large charge Q passes between the electrodes when $\theta=0$ at time $t=0$ and is followed by no other charge, the reading of the instrument at all subsequent times is given by

$$\theta = Q \cdot f(t) \quad \dots \dots \dots (3.0)$$

It is assumed that $f(t)$ is linear, both in the sense that $f(t)$ is independent of Q and that the effect of two successive changes is the sum of their individual effects ; further that

$$f(t)=0 \text{ for } t<0 \text{ and for } t=\infty \quad \dots \dots \dots (3.2)$$

The bar over θ or θ^2 denoted an average of values taken at instants during the passage of the electrons selected at random. It is assumed that if θ is read continuously (*i. e.*, at all instants) over a long time T ,

the effect of " mean operator " $\frac{1}{T} \int_T dt$ is to generate that average.

4. Let

$$F(j\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} f(t) e^{j\omega t} dt, \quad \dots \dots \dots (4.1)$$

so that F is the Fourier transform of f . Then, since $f(t)$ is real and zero for $t<0$, in virtue of a fundamental proposition of Fourier transforms,

$$\int_0^{\infty} f^2(t) dt = \int_{-\infty}^{+\infty} |f(t)|^2 dt = \int_{-\infty}^{+\infty} |F(j\omega)|^2 d\omega. \quad \dots \dots (4.2)$$

If I is the mean current passing between the electrodes,

$$I = N\epsilon, \quad \dots \dots \dots (4.3)$$

and we can write (1.1)

$$\bar{\theta}^2 - (\bar{\theta})^2 = I\epsilon \int_{-\infty}^{+\infty} |F(j\omega)|^2 d\omega. \quad \dots \dots \dots (4.4)$$

5. An examination of the proof of Campbell's theorem * shows that (1.1) is true even if the electrons occupy a finite transit time τ_0 in passing between the electrodes. But if the transit time is finite, $f(t)$, characteristic of an instantaneous passage, must be replaced by some other function $f_i(t)$. The following argument, which we owe to Mr. D. O. North of The Radio Corporation of America, enables $f_i(t)$ to be calculated.

A charge Q , starting at $t=0$ and having a transit time τ_0 , is equivalent to a current pulse $Q \cdot i(\tau)$, where

$$i(\tau)=0 \text{ for } \tau<0 \text{ or } >\tau_0, \quad . \quad . \quad . \quad . \quad . \quad (5.1)$$

$$\int_0^{\tau_0} i(\tau) d\tau=1.$$

This pulse during the infinitesimal interval $d\tau$ about $t=\tau$ is equivalent to the instantaneous passage of a charge $Q \cdot i(\tau)d\tau$. For this charge the function corresponding to $f(t)$ is $f(t-\tau)$, for its passage dates from $t=\tau$, not from $t=0$. Consequently the whole effect of the charge is

$$\theta=Q \cdot f_i(t)=\int_0^{\tau_0} Q \cdot i(\tau) \cdot f(t-\tau)d\tau, \quad . \quad . \quad . \quad . \quad . \quad (5.2)$$

or, in view of (3.2), (5.1),

$$f_i(t)=\int_{-\infty}^{+\infty} i(\tau) \cdot f(t-\tau)d\tau. \quad . \quad . \quad . \quad . \quad . \quad (5.3)$$

Let $g(j\omega)$ be the Fourier transform of $i(\tau)$ and $F_i(j\omega)$ the Fourier transform of $f_i(t)$. Then, since $F(j\omega)$ is the Fourier transform of $f(t)$, we have from (5.3), in virtue of a general property of Fourier transforms,

$$F_i(j\omega)=\sqrt{2\pi}g(j\omega)F(j\omega), \quad . \quad . \quad . \quad . \quad . \quad (5.4)$$

so that (4.4) becomes

$$\overline{\theta^2}-(\overline{\theta})^2=2\pi I\epsilon \int_{-\infty}^{+\infty} |g(j\omega)F(j\omega)|^2 d\omega=2\pi I\epsilon \int_{-\infty}^{+\infty} |g(j\omega)|^2 \cdot |F(j\omega)|^2 d\omega. \quad . \quad . \quad . \quad . \quad . \quad (5.5)$$

6. We turn now to Schottky's theorem. The basic conception here is that we are concerned only with sinusoidal oscillations whose frequency lies within a band of width δf , within which everything is independent of frequency. The proposition states that then we may replace the randomly arriving electrons by a sinusoidal current whose frequency lies within this range and whose amplitude i is given by (1.2).

The best derivation—or, it might be better to say, interpretation—of this proposition is that of Ballantine †; since he does not give all the

* The latest proof is that of J. M. Whittaker, Proc. Camb. Phil. Soc. xxxiii. p. 456 (1937).

† Stuart Ballantine, J. Franklin Soc. civi. p. 159 (1928).

steps in his argument, we reproduce it here slightly expanded. The arrival of an electron is again regarded as equivalent to a current pulse $\epsilon \cdot i(T)$. If this pulse is resolved into sinusoidal oscillations, the amplitude of the oscillation of frequency $\omega/2\pi$ is $\epsilon \cdot g(j\omega)$. During the long time T of paragraph 3, the number of those trains of oscillations generated is NT ; since the electrons arrive at random instants, the phases of these trains will be distributed at random. If NT is so great that statistically most probable values are the actual values, then, by Rayleigh's theorem*, these trains of frequency $\omega/2\pi$ are equivalent, in all matters that involve only amplitudes and not phases, to a single train, whose amplitude i_ω is given by

$$i_\omega = \sqrt{NT} \cdot \epsilon \cdot g(j\omega), \quad \dots \dots \dots (6.1)$$

and all the trains of all frequencies are equivalent in the same sense to a current $I(t)$, where the Fourier transform of $I(t)$ is i_ω . Hence, by the property involved in (4.2),

$$\int_{-\infty}^{+\infty} I^2(t) dt = \int_{-\infty}^{+\infty} NT \cdot \epsilon^2 \cdot |g(j\omega)|^2 d\omega. \quad \dots \dots \dots (6.2)$$

When $T \rightarrow \infty$, the operator $\frac{1}{T} \int_{-\infty}^{+\infty} dt$ is equivalent to the bar of paragraph 3. Hence, using (4.3),

$$\overline{I^2(t)} = \frac{1}{T} \int_{-\infty}^{+\infty} I^2(t) dt = I \epsilon \int_{-\infty}^{+\infty} |g(j\omega)|^2 d\omega. \quad \dots \dots \dots (6.3)$$

This proposition may be called Ballantine's theorem. The proof given of it is of about the same degree of rigidity as Whittaker's proof of Campbell's theorem, and is subject to difficulties of the same kind as Rowland† discussed in connexion with the latter. But if we take Ballantine's theorem as fundamental to one method of treatment as we have taken Campbell's to the other, these difficulties need not concern us.

7. If i^2 is identified with $\overline{I^2(t)}$, (1.2) is the form that (6.3) takes when τ_0 , the duration of each current pulse, is zero. For then

$$\begin{aligned} \epsilon \cdot g(j\omega) &= \frac{\epsilon}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} i(\tau) e^{j\omega\tau} d\tau = \frac{\epsilon}{\sqrt{2\pi}} \int_0^{\tau_0} i(t) e^{j\omega t} dt \\ &= \frac{\epsilon}{\sqrt{2\pi}} e^{j\omega\xi\tau_0} \cdot \int_0^{\tau_0} i(t) dt \quad (0 < \xi < 1). \quad \dots \dots \dots (7.1) \end{aligned}$$

or, remembering (5.1),

$$\doteq \frac{\epsilon}{\sqrt{2\pi}} (1 + j\omega\xi\tau_0), \text{ if } \omega\tau_0 \ll 1. \quad \dots \dots \dots (7.2)$$

* Lord Rayleigh, 'Theory of Sound,' paragraph 42 *a* (1894).

† E. N. Rowland, Proc. Camb. Phil. Soc. xxxii. p. 580 (1936).

Proceeding to the limit $\tau_0 \rightarrow 0$

$$g(j\omega) = \frac{1}{\sqrt{2\pi}} \quad \dots \quad (7.3)$$

Further, since $g(j\omega)$ is an even function of ω , we can replace $\int_{-\infty}^{+\infty}$ by $2 \int_0^{\infty}$; and if $|g(j\omega)|^2$ is constant, δf means $\frac{1}{2\pi} \int_0^{\infty} d\omega$. (6.3) therefore reduces to (1.2).

8. A certain similarity between (5.5) and (6.3) is immediately apparent. But there is no term in (6.3) corresponding to $\bar{(\theta)}^2$ in (5.5); this difference requires some consideration.

In order that equilibrium should be established between the components of random phases, interaction between them must continue over many periods of the oscillations. (6.3) cannot therefore be true if $g(j\omega)$ is finite for values of ω comparable with $1/T$, where T is the time for which the experiment lasts. In actual experiments the apparatus is always arranged so that the effect of the constant component of the current is eliminated, that is to say, so that components of the current with sufficiently small ω produce no effect at the output terminals where the noise is measured. If so, it does not matter what $g(j\omega)$ is for small values of ω , and (6.3) can still be used. In these circumstances $\bar{\theta}$ in (5.5) will be zero, so that the difference between (5.5) and (6.3) is only apparent.

If, on the other hand, the constant component were not eliminated by the apparatus, we should have to divide the Fourier transform of the current pulse into two parts, namely g_1 , applicable to frequencies large compared with $1/T$ and zero for small frequencies, and g_2 applicable to the smaller frequencies. (6.3) would then be true if g_1 were substituted for g and I_1 for I , where I_1 is a current producing the same effect as the components of higher frequencies. If I_2 were the current producing the same effect as the lower frequencies, the total effect would be the sum of those due to I_1 and I_2 , which would have to be calculated separately. If I were still the current equivalent to all components, $I = I_1 + I_2$. It might be difficult to calculate I_2 accurately; but it is easy to see that I_2 would be nearly I . We should have then

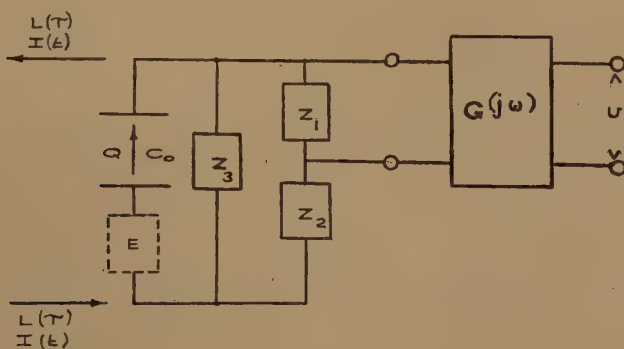
$$\overline{I_1^2} = \overline{(I - I_2)^2} = \overline{I^2} - \overline{(I_2)^2} = \int_{-\infty}^{+\infty} NT\epsilon^2 |g_1(j\omega)|^2 d\omega, \quad \dots \quad (8.1)$$

which is similar to (5.5).

It may be pointed out that no similar difficulties enter if we start from (1.1). The proof of (5.5) is valid, even if the constant component of θ is taken into account, so long as T is large compared with the largest

value of t for which $f(t)$ is appreciably different from zero. Thus, if we prove that the two methods are equivalent, we shall prove that (8.1) is true, although it has not been proved to follow from Ballantine's argument.

9. (5.5) and (6.3) are then mathematically similar; we have now to show that they are physically identical. For this purpose we must consider the circuit connected to the diode. We may assume that the shot noise actually measured is a voltage; the most general circuit that we have to consider is that shown in the figure. C_0 is the interelectrode capacity of the diode, Z_3 is an impedance connected between the electrodes. If Z_3 includes a capacity directly between the electrodes, enquiry shows that it makes no difference whether we do or do not take this capacity out of Z_3 and include it in C_0 . Z_1 and Z_2 are two other impedances forming a potentiometer across Z_3 . The voltage across Z_1



is applied to the input of an amplifier of gain G , whose input admittance (if any) is included in Z_1 , so that the input impedance of the amplifier may be taken as infinite. G is the ratio of output to input voltage, taking into account change of phase, so that $|G|$ is the amplitude gain. v_r is the voltage across Z_r ; the noise voltage v is taken across the output of the amplifier. The arrows show where Q passes across to paragraph 3, and where $i(\tau)$ and $I(t)$ enter and leave according to paragraph 6. The significance of E will appear later.

In virtue of the considerations of paragraph 8, there must be some feature eliminating the effect of direct current. In practice $G(0)=0$, but in principle the direct component of v_1 or v_3 might be compensated from some other source.

It will be convenient to write Z for the impedance of the (Z_1, Z_2, Z_3), so that

$$\frac{1}{Z} = \frac{1}{Z_3} + \frac{1}{Z_1 + Z_2}, \quad \dots \quad (9.1)$$

and Z' for the impedance of Z and C_0 in parallel,

$$Z'(j\omega) = \frac{Z(j\omega)}{1 + j\omega C_0 Z(j\omega)} \tag{9.2}$$

10. In proving our proposition, we shall employ Heaviside analysis. The Z 's and G are then at the outset operators that are functions of the Heaviside operator p .

If we start from (6.2), we have

$$v_3(t) = Z'(p)I(t) \tag{10.1}$$

and
$$v(t) = \frac{Z_1(p)}{Z_1(p) + Z_2(p)} G(p)v_3(t). \tag{10.2}$$

But $\sqrt{NT} \cdot \epsilon \cdot g(j\omega)$ is the Fourier transform of $I(t)$, and

$$Z(p)e^{pt} = Z(\lambda)e^{\lambda t}. \tag{10.3}$$

Hence
$$v_3(t) = Z'(p) \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \sqrt{NT} \cdot \epsilon \cdot g(j\omega) e^{-j\omega t} d\omega \tag{10.4}$$

$$= \frac{\sqrt{NT} \cdot \epsilon}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} g(j\omega) Z'(-j\omega) e^{-j\omega t} d\omega. \tag{10.5}$$

Hence, in virtue of the propositions involved in the passage from (4.1) to (4.2) and from (6.2) to (6.3)

$$|\overline{v_3(t)}|^2 = I\epsilon \int_{-\infty}^{+\infty} |g(j\omega)|^2 \cdot |Z'(-j\omega)|^2 d\omega \tag{10.6}$$

$$|\overline{v(t)}|^2 = I\epsilon \int_{-\infty}^{+\infty} |g(j\omega)|^2 |Z'(-j\omega)|^2 \left| \frac{Z_1(-j\omega)}{Z_1(-j\omega) + Z_2(-j\omega)} \right|^2 |G(-j\omega)|^2 d\omega. \tag{10.7}$$

11. If we arrive at (10.6) from any other starting point, we shall arrive also at (10.7). Hence it is sufficient to show that we arrive at (10.6) if we start from (5.5). For that purpose we must show that, when $Qf(t)$ is $v_3(t)$,

$$2\pi |F(j\omega)|^2 = |Z'(-j\omega)|^2 \tag{11.1}$$

Since, in arriving at (5.5), no assumption was made about the physical meaning of ω —as it was in arriving at (6.3)—we can now identify the ω of paragraph 5 with that of paragraph 6.

We first show that the current $i_z(\tau)$ flowing in Z is the same whether it is due to a current pulse $Q \cdot i(\tau)$, for which $\tau_0 \rightarrow 0$, flowing as shown in fig. 1, or due to the sudden transference of a charge Q as shown in that figure, so long as the total charge conveyed is the same in both cases.

Since $g(j\omega)$ is the Fourier transform of $i(\tau)$, and Z' is given by (9.2),

$$i_z(\tau) = \frac{Z'(p)}{Z(p)} Q i(\tau) = \frac{Q}{1 + C_0 p Z(p)} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} g(j\omega) e^{-j\omega \tau} d\omega. \tag{11.2}$$

But from (7.3) when $\tau \rightarrow 0$

$$g(j\omega) \doteq \frac{1}{\sqrt{2\pi}} (1 + j\omega \xi \tau_0). \quad (11.3)$$

Hence

$$i_z(\tau) = \frac{Q}{2\pi} \int_{-\infty}^{+\infty} \frac{1 + j\omega \xi \tau_0}{1 - j\omega C_0 Z(-j\omega)} e^{-j\omega \tau} d\omega, \quad (11.4)$$

or, changing the variable and inserting λ in numerator and denominator,

$$= \frac{Q}{2\pi j} \int_{-j\infty}^{+j\infty} \frac{(1 - \lambda \xi \tau_0) \lambda}{1 + \lambda C_0 Z(\lambda)} \frac{e^{\lambda \tau}}{\lambda} d\lambda \quad (11.5)$$

Using the known formula, in which H is the Heaviside unit function,

$$Q(p)H(t) = \frac{1}{2\pi j} \int_{-j\infty}^{+j\infty} \frac{Q(\lambda)}{\lambda} e^{\lambda t} d\lambda. \quad (11.6)$$

we have

$$i_z(\tau) = \frac{Q(p - \xi \tau_0 p^2)}{1 + C_0 p \cdot Z(p)} H(\tau), \quad (11.7)$$

On the other hand, according to a proposition often used in Heaviside analysis, the sudden transfer of the charge Q across C_0 is equivalent to the sudden introduction at E , in series with C_0 , of a source of E.M.F. Q/C_0 and of zero internal impedance. The current in Z thereby produced is

$$\begin{aligned} i_z(\tau) &= \frac{Q}{C_0} \left/ \left(\frac{1}{C_0 p} + Z(p) \right) \right. \cdot H(\tau) \\ &= \frac{Qp}{1 + C_0 p \cdot Z(p)} \cdot H(\tau). \end{aligned} \quad (11.8)$$

(11.7) and (11.8) become identical as $\tau_0 \rightarrow 0$. The first part of the proof is thereby complete.

But, if $Q \cdot f(t)$ is identified with $v_s(t)$, and τ is replaced by t ,

$$Qf(t) = Z(p)i_z(t), \quad (11.9)$$

or, in virtue of (11.4) when $\tau_0 \rightarrow 0$,

$$\begin{aligned} f(t) &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{Z(-j\omega)}{1 - j\omega C_0 \cdot Z(-j\omega)} e^{-j\omega t} d\omega \\ &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} Z'(-j\omega) e^{-j\omega t} d\omega. \end{aligned} \quad (11.10)$$

But

$$f(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} F(j\omega) e^{-j\omega t} d\omega, \quad (11.11)$$

whence

$$F(j\omega) = \frac{1}{\sqrt{2\pi}} Z'(-j\omega). \quad (11.12)$$

(11.1) follows immediately and the equivalence of the two methods is proved.

However, an assumption has been made implicitly which must be examined. (11.6) is not true if $Q(\lambda)/\lambda$ has a pole on the imaginary axis. The presence of λ in the denominator does not prove that there is a pole at $\lambda=0$, for $Q(\lambda)$ has a factor λ . If $Q(\lambda)/\lambda$ had a pole at $j\omega_0$, $Q(p)H(t)$ would have a term in $e^{j\omega_0 t}$ with a non-zero coefficient. That would mean that the system was capable of undamped, and therefore indefinitely prolonged, oscillations. We have already excluded this explicitly in § 3 by providing that $f(\infty)=0$; the corresponding provision must be made, though it is not so easy to state explicitly, if we start from (1.2).

12. Some general remarks may be added. It is sometimes said that the two methods differ because Campbell's theorem concerns charges and Ballantine's concerns currents. But that is not a real difference; for, unless the calculations are in error, no difference can arise on classical principles by substituting a current pulse for the passage of a charge, or vice versa. The real difference in the two methods is that, whereas the use of Ballantine's theorem involves the Fourier analysis of the disturbances before their effect on the circuit is considered, the use of Campbell's theorem involves the Fourier analysis of the effect they produce on the circuit. Both methods necessarily involve a "correspondence theorem," that propositions known to be valid when large numbers of atomic events are concerned are valid also when single events are concerned. In Campbell's theorem this assumption is explicit and takes the form that $f(t)$, determined by classical principles, is applicable to a single electron. In Ballantine's theorem it is better concealed, but is involved in the proposition that Fourier analysis may be applied to a current pulse consisting of a single electron. The conception of randomness, essential to the discussion, is introduced in either case by the characteristic proposition (1.2) or (6.3); an examination of the "proofs" shows that in each case "randomness" is defined by the postulate that a random deviation from the mean is as likely to be positive as negative; but since neither proof is entirely convincing, these propositions themselves should be regarded as defining "random." The question then arises whether the definitions are physically identical; the proof that the propositions must always lead to the same experimental result answers this question affirmatively.

XLIX. *The Numerical Solution of a Type of Equation.*

By H. C. PLUMMER *.

[Received July 16, 1941.]

1. THE general type of equation to be considered is

$$\tan x = xf(x), \quad \dots \dots \dots (1)$$

where $f(x)$ is a single-valued function. In general $f(x) = f(-x)$, so that $x = 0$ and $x = \pm x_n$ are roots. In any interval between $x = (n \pm \frac{1}{2})\pi$, $\tan x$ increases from $-\infty$ to $+\infty$, and in general there is at least one root x_n in the interval.

2. The simplest example is furnished by the familiar equation †

$$\tan x = x, \quad f(x) = 1. \quad \dots \dots \dots (2)$$

The intersections of $y = \tan x$, $y = x$ show that the roots are of the form ($n = 1, 2, \dots$),

$$x_n = (n + \frac{1}{2})\pi - \beta_n, \quad \dots \dots \dots (3)$$

where β_n is a positive angle diminishing as x increases. Thus

$$\cot \beta_n = (n + \frac{1}{2})\pi - \beta_n$$

$$\text{and} \quad \frac{2}{\sin 2\beta_n} = \cot \beta_n + \tan \beta_n = (n + \frac{1}{2})\pi + (\tan \beta_n - \beta_n). \quad \dots \dots \dots (4)$$

Hence a first approximation to the value of β_n is given by

$$\beta_n = \frac{1}{2} \arcsin \frac{4}{(2n+1)\pi}, \quad \dots \dots \dots (5)$$

which can afterwards be corrected by (4). But comparison of (5) with the Euler-Rayleigh expansion shows that the error is of the order $[3\pi^5(n + \frac{1}{2})^5]^{-1}$, so that β_1 (in radians) is in defect of the true value by about 1 in the fourth place, β_2 by 1 in the fifth place, β_4 by 1 in the sixth place and β_6 by less than 1 in the seventh place of decimals. As four places are generally enough the needful accuracy is given by (5) without correction except perhaps when $n = 1$. Then (5) gives $x_1 = 4.4933$, whereas the first positive root is more accurately 4.4934095.

* Communicated by the Author.

† Rayleigh's 'Sound,' i. p. 278 (1877).

3. When $f(x)=m$, a constant, the distribution of the roots is similar. If $m>1$, there are three roots in the interval between $x=\pm\frac{1}{2}\pi$. If $m<0$ the positive roots are $x=(n-\frac{1}{2})\pi+\beta_n$, where β_n is a positive angle decreasing as $n(=1, 2, \dots)$ increases. There are particular cases which deserve notice.

$$(a) \quad m=\frac{1}{2}, \quad \tan x=\frac{1}{2}x. \quad \dots \dots \dots (6)$$

The distribution of the positive roots is given by (3), where

$$\cot \beta_n = \frac{1}{2}(n+\frac{1}{2})\pi - \frac{1}{2}\beta_n, \quad n=1, 2, \dots$$

$$\text{or} \quad \cot \beta_n + \tan \frac{1}{2}\beta_n = \frac{1}{\sin \beta_n} = \frac{1}{2}(n+\frac{1}{2})\pi + (\tan \frac{1}{2}\beta_n - \frac{1}{2}\beta_n). \quad \dots (7)$$

The first approximation, which gives the higher roots with increasing accuracy, is therefore

$$\beta_n = \arcsin \frac{4}{(2n+1)\pi}. \quad \dots \dots \dots (8)$$

$$(b) \quad m=\frac{1}{4}, \quad \tan x=\frac{1}{4}x. \quad \dots \dots \dots (9)$$

The roots are again given by (3) when

$$\cot \beta_n = \frac{1}{4}(n+\frac{1}{2})\pi - \frac{1}{4}\beta_n, \quad n=1, 2, \dots$$

$$\text{or} \quad \cot \beta_n + \frac{1}{2} \tan \frac{1}{2}\beta_n = \frac{1}{2} \cot \frac{1}{2}\beta_n = \frac{1}{4}(n+\frac{1}{2})\pi + \frac{1}{2}(\tan \frac{1}{2}\beta_n - \frac{1}{2}\beta_n). \quad \dots (10)$$

The first approximation is therefore

$$\beta_n = 2 \arctan \frac{4}{(2n+1)\pi}. \quad \dots \dots \dots (11)$$

$$(c) \quad m=-1, \quad \tan x+x=0. \quad \dots \dots \dots (12)$$

Intersections of the line $x+y=0$ with $y=\tan x$ show that the positive roots are of the form

$$x_n = (n-\frac{1}{2})\pi + \beta_n, \quad n=1, 2, \dots \quad \dots \dots (13)$$

$$\text{where} \quad \cot \beta_n = (n-\frac{1}{2})\pi + \beta_n,$$

$$\text{or} \quad \cot \beta_n - \tan \beta_n = 2 \cot 2\beta_n = (n-\frac{1}{2})\pi - (\tan \beta_n - \beta_n). \quad \dots (14)$$

Hence the first approximation is

$$\beta_n = \frac{1}{2} \arctan \frac{4}{(2n-1)\pi}. \quad \dots \dots \dots (15)$$

In all these cases the higher roots beyond a certain stage will be given by the first approximation. Better values of the earlier roots can be found by using the corresponding exact forms of equation or by more direct methods.

4. The next equation to be considered is

$$\tan x = \frac{x}{1-x^2}, \quad \dots \dots \dots (16)$$

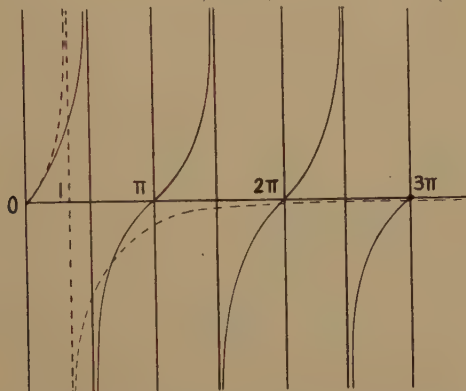
which relates to electrical waves on a sphere *. From the intersections of the curves

$$y = \tan x, \quad y(1-x^2) = x,$$

the asymptotes of the latter being $x = \pm 1$ and $y = 0$ (fig. 1), the distribution of the roots is clear. They are

$$x_n = n\pi - \beta_n, \quad n = 1, 2, \dots \dots \dots (17)$$

Fig. 1.



where β_n is positive and decreasing as n increases. Thus

$$\tan \beta_n = \frac{n\pi - \beta_n}{(n\pi - \beta_n)^2 - 1},$$

or $\cot \beta_n + \tan \beta_n = \frac{2}{\sin 2\beta_n} = n\pi - \frac{1}{n\pi} + (\tan \beta_n - \beta_n) - \frac{\beta_n}{n\pi(n\pi - \beta_n)} \quad (18)$

Therefore the first approximation is

$$\beta_n = \frac{1}{2} \arcsin \frac{2n\pi}{n^2\pi^2 - 1}, \quad \dots \dots \dots (19)$$

which gives for the first five values

$$(\beta_1') \ 0.3935, \ (\beta_2) \ 0.1664, \ (\beta_3) \ 0.1081, \ (\beta_4) \ 0.0804, \ (\beta_5) \ 0.0641.$$

To this degree of accuracy (19) is adequate from $n=2$ onward. As $\beta_n \rightarrow 1/n\pi$, a second approximation to (18), is

$$2 \operatorname{cosec} 2\beta_n = n\pi - \frac{1}{n\pi} - \frac{2}{3n^3\pi^3}, \quad \dots \dots \dots (19a)$$

* J. J. Thomson, 'Recent Researches in Electricity and Magnetism,' p. 373 (1893).

which gives $\beta_1''=0.3975$. A more accurate value, found on verification, is $\beta_1=0.3980$, or $x_1=2.7436$. J. J. Thomson's value is 2.744.

Thus, for the first five roots of (17),

$$\frac{x_1}{\pi}=0.8733, \quad \frac{x_2}{2\pi}=0.9735, \quad \frac{x_3}{3\pi}=0.9885, \quad \frac{x_4}{4\pi}=0.9936, \quad \frac{x_5}{5\pi}=0.9959,$$

and the first three are in fair agreement with values given by Silberstein *. His fourth root seems to be appreciably in error, and he overrates the approximation of x_n to $n\pi$. For the relative divergence is of the order $(n\pi)^{-2}$ and must amount to 10^{-3} when $n=10$.

5. Sound waves in a sphere give rise to the equation †

$$\tan x = \frac{2x}{2-x^2}. \quad . \quad . \quad . \quad . \quad . \quad (20)$$

The distribution of the roots being given by

$$y = \tan x, \quad y(2-x^2) = 2x,$$

of which the latter has the asymptotes $x = \pm\sqrt{2}$ and $y=0$, it shares the same character as those shown in fig. 1. Hence it is possible again to write

$$x_n = n\pi - \beta_n, \quad n=1, 2, \dots \quad . \quad . \quad . \quad . \quad . \quad (17)$$

where

$$\tan \beta_n = \frac{2(n\pi - \beta_n)}{(n\pi - \beta_n)^2 - 2},$$

$$\cot \beta_n + \tan \frac{1}{2}\beta_n = \frac{1}{2}(n\pi - \beta_n) - \frac{1}{n\pi - \beta_n} + \tan \frac{1}{2}\beta_n,$$

$$\frac{1}{\sin \beta_n} = \frac{1}{2}n\pi - \frac{1}{n\pi} - \frac{\beta_n}{n\pi(n\pi - \beta_n)} + (\tan \frac{1}{2}\beta_n - \frac{1}{2}\beta_n). \quad . \quad . \quad (21)$$

The first approximation is therefore

$$\beta_n = \arcsin \frac{2n\pi}{n^2\pi^2 - 2}, \quad . \quad . \quad . \quad . \quad . \quad (22)$$

to which correspond

$$(\beta_1') 0.9243, (\beta_2') 0.3419, (\beta_3') 0.2188, (\beta_4) 0.1619, (\beta_5) 0.1287.$$

On verification an adjustment to $\beta_3=0.2189$ is suggested; otherwise (22) is adequate for the higher roots. Since $\beta_n \rightarrow 2/n\pi$, a second approximation to (21) gives

$$\operatorname{cosec} \beta_n = \frac{1}{2}n\pi - \frac{1}{n\pi} - \frac{5}{3n^3\pi^3}. \quad . \quad . \quad . \quad . \quad (22 a)$$

* Bell's 'Mathematical Tables' (1922).

† Rayleigh's 'Sound,' ii. p. 232 (1878).

This gives $\beta_2=0.3428$, which is verified, while $\beta_1''=0.9866$, which is still only approximate. It leads easily, when trigonometrical tables* with the argument expressed in radians are used, to the value $\beta_1=1.0600$. Thus the first five roots are $x=n\pi-\beta_n$, where

$$\beta_1=1.0600, \beta_2=0.3428, \beta_3=0.2189, \beta_4=0.1619, \beta_5=0.1287,$$

and additional roots are given with the same accuracy by (22). Otherwise expressed,

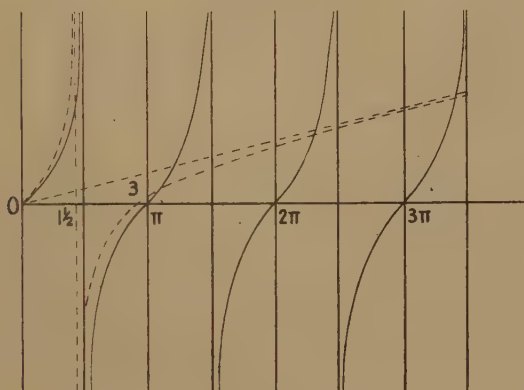
$$x_1=119^\circ.26, \quad x_2=340^\circ.36$$

in almost complete agreement with Rayleigh's values †.

6. An equation related to (20) and taken from the same source is

$$\tan x = \frac{x(x^2-9)}{4x^2-9}, \quad \dots \dots \dots (23)$$

Fig. 2.



which approximates to (9) when x is large. The curves

$$y=\tan x, \quad y(4x^2-9)=x(x^2-9)$$

have the form shown in fig. 2.

The intersecting curve has the three asymptotes $2x=\pm 3$ and $x=4y$, and passes through the point (3, 0). Its form shows clearly that

$$x_1=\pi+\beta_1, \quad x_n=(n+\frac{1}{2})\pi-\beta_n \quad \dots \dots \dots (24)$$

($n=2, 3, \dots$), where β_n is a small positive angle. The first root must be calculated separately.

* Such as Milne-Thomson and Comrie, 'Standard Four-figure Mathematical Tables.'

† Rayleigh's 'Sound,' ii. p. 232.

In this case

$$\begin{aligned}\tan \beta_1 \{4\pi^2 - 9 + 8\pi\beta_1 + 4\beta_1^2\} &= \pi(\pi^2 - 9) + 3(\pi^2 - 3)\beta_1 + 3\pi\beta_1^2 + \beta_1^3 \\ \pi(\pi^2 - 9) - \pi^2\beta_1 - 5\pi\beta_1^2 &= (\tan \beta_1 - \beta_1) \{4\pi^2 - 9 + 8\pi\beta_1 + 4\beta_1^2\} + 3\beta_1^3.\end{aligned}$$

The right-hand side being denoted by B, of the order of β_1^3 , the positive root is

$$\begin{aligned}\beta_1 &= -0.1\pi + \sqrt{\{0.21\pi^2 - 1.80 - 0.2B/\pi\}} \\ &= -0.31416 + \sqrt{\{0.27262 - B/5\pi\}}\end{aligned}$$

or $\beta_1 < 0.208$ for the first approximation, when B is ignored. When β_1^5 is neglected in B, and this value of β_1 is taken,

$$B = \frac{4}{3} \beta_1^3 (\pi + \beta_1)^2, \quad B/5\pi = 0.00853,$$

and the next value of β_1 is 0.1993. Hence it is easy to arrive at $\beta_1 = 0.2006$ or $x_1 = 3.3422$, which is Rayleigh's value, though it is a little beyond the scope of four-figure tables to ensure the last figure.

For the other roots, according to (24) and (23),

$$\begin{aligned}\cot \frac{1}{2}\beta_n &= \tan \frac{1}{2}\beta_n + 2 \cot \beta_n = \tan \frac{1}{2}\beta_n + \frac{2x_n(x_n^2 - 9)}{4x_n^2 - 9} \\ &= (\tan \frac{1}{2}\beta_n + \frac{1}{2}x_n - \frac{27x_n}{2(4x_n^2 - 9)}) \\ &= (\tan \frac{1}{2}\beta_n - \frac{1}{2}\beta_n) + \frac{1}{2}(n + \frac{1}{2})\pi - \frac{27\{(n + \frac{1}{2})\pi - \beta_n\}}{2[4\{(n + \frac{1}{2})\pi - \beta_n\}^2 - 9]} \\ &= \frac{1}{2}(n + \frac{1}{2})\pi - \frac{27(n + \frac{1}{2})\pi}{2\{4(n + \frac{1}{2})^2\pi^2 - 9\}} + (\tan \frac{1}{2}\beta_n - \frac{1}{2}\beta_n) \\ &\quad - \frac{27\beta_n[4(n + \frac{1}{2})\pi\{(n + \frac{1}{2})\pi - \beta_n\} + 9]}{2[4(n + \frac{1}{2})^2\pi^2 - 9][\{4(n + \frac{1}{2})\pi - \beta_n\}^2 - 9]} \dots \dots \dots (25)\end{aligned}$$

Hence the first approximation is

$$\beta_n = 2 \arctan \frac{\{4(n + \frac{1}{2})^2\pi^2 - 9\}}{2(n + \frac{1}{2})\pi\{(n + \frac{1}{2})^2\pi^2 - 9\}}, \dots \dots \dots (26)$$

from which the first four values are

$$(\beta_2') \ 0.5594, (\beta_3') \ 0.3811, (\beta_4') \ 0.2908, (\beta_5) \ 0.2359.$$

The last is approximately correct, and beyond this point (26) will give β_n with increasing accuracy. For the rest the process of verification suggests the needed adjustments, namely

$$\beta_2 = 0.5640, \beta_3 = 0.3817, \beta_4 = 0.2910.$$

By (24) the first five roots of (23) have thus been found.

7. The equations

$$\cos x \cosh x = \pm 1, \dots \dots \dots (27)$$

which are connected with the lateral vibrations of a rod * (free-free or clamped-free), can be written

$$\tan \frac{1}{2}x = \pm \tanh \frac{1}{2}x, \quad \tan \frac{1}{2}x = \pm \coth \frac{1}{2}x.$$

These are of the type considered above. Perhaps the nature of the roots is most simply shown by the intersection of the curve

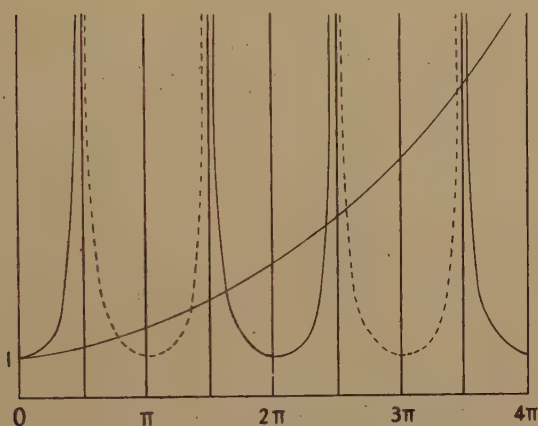
$$y = \cosh x, \quad y = \pm \sec x,$$

a catenary † and a secant curve (with its reversal) (fig. 3).

It is clear that the roots are $(x=1, 2, 3, \dots)$

$$\left. \begin{aligned} +1, \quad x_n &= (n + \frac{1}{2})\pi - (-1)^n \beta_n \\ -1, \quad x_n' &= (n - \frac{1}{2})\pi - (-1)^n \alpha_n \end{aligned} \right\}, \quad \dots \quad (28)$$

Fig. 3.



where $\alpha_1, \alpha_2, \beta_1, \beta_2, \alpha_3, \alpha_4, \dots$ is a decreasing sequence of positive angles rapidly converging to 0. Thus

$$\begin{aligned} \cos x_n &= \sin \beta_n, \quad \cos x_n' = -\sin \alpha_n \\ 0 &= \sin \beta_n (e^{2x} + 1) - 2e^{x_n} = \sin \alpha_n (e^{2x_n'} + 1) - 2e^{x_n'} \\ e^{x_n} &= \frac{1 + \cos \beta_n}{\sin \beta_n} = \cot \frac{1}{2}\beta_n, \quad e^{x_n'} = \cot \frac{1}{2}\alpha_n, \end{aligned}$$

positive roots only being considered. Hence (in common logs)

$$\begin{aligned} -\log \tan \frac{1}{2}\beta_n &= \{(n + \frac{1}{2})\pi - (-1)^n \beta_n\} \log e \\ -\log \tan \frac{1}{2}\alpha_n &= \{(n - \frac{1}{2})\pi - (-1)^n \alpha_n\} \log e \end{aligned}$$

* Rayleigh's 'Sound,' i. p. 222.

† The ordinate of the catenary increases far too rapidly to be represented on anything like a true scale in the figure.

or

$$\begin{aligned}\log \beta_n &= \log 2 - (n + \tfrac{1}{2})\pi \log e + (-1)^n \beta_n \log e - \log (\tan \tfrac{1}{2} \beta_n / \tfrac{1}{2} \beta_n) \\ &= (2n + 1) \bar{1} \cdot 3178117 + 0 \cdot 3010300 + (-1)^n \beta_n M - \log (\tan \tfrac{1}{2} \beta_n / \tfrac{1}{2} \beta_n) \\ \log \alpha_n &= (2n - 1) \bar{1} \cdot 3178117 + 0 \cdot 3010300 + (-1)^n \alpha_n M - \log (\tan \tfrac{1}{2} \alpha_n / \tfrac{1}{2} \alpha_n).\end{aligned}\quad \dots \quad (29)$$

The last term is negligible for β_2 , α_3 and beyond. If 10^{-7} is set as the limit of α or β , this limit is passed by β_5 , α_6 . For α_1 the more convenient form is

$$\log \tan \tfrac{1}{2} \alpha_1 = \bar{1} \cdot 3178117 - M \alpha_1, \quad \dots \quad (30)$$

which is satisfied by Rayleigh's solution

$$\alpha_1 = 0 \cdot 3043077 = 17^\circ 26' 8''.$$

The other values are easily found, namely

$$\begin{array}{ll}\beta_1 = 0 \cdot 0176518, & \alpha_2 = 0 \cdot 0182978 \\ \beta_2 = 0 \cdot 0007770, & \alpha_3 = 0 \cdot 0007758 \\ \beta_3 = 0 \cdot 0000336, & \alpha_4 = 0 \cdot 0000336 \\ \beta_4 = 0 \cdot 0000014, & \alpha_5 = 0 \cdot 0000014,\end{array}$$

which give all the roots which differ by as much as 10^{-7} from $(n + \frac{1}{2})\pi$. The roots are given correctly in 'Sound,' 2nd ed., but in the first edition α_{n+1} is wrongly identified with β_n . The oversight was noted by Greenhill, who has also given other interesting application of the same period equations*.

8. Two points emerge from the consideration of these numerical solutions. The first is the value of the graphical representation of the roots in all these cases. The oversight just noticed would have been impossible if such a diagram as fig. 3 had been roughly sketched, for Rayleigh's arithmetic was extremely accurate. The general location of the roots is often more useful and probably more instructive than any calculated values. The second point is that by suitable choice of a simple function an effective approximation can be found which becomes a solution accurate enough in itself for the roots beyond a certain stage, though in the case of the smaller roots it serves only as a useful starting point in the search for more accurate values.

* Mess. of Maths. xvi. p. 115 (1886).

INDEX TO VOL. XXXII.

- ADMITTANCE** method for obtaining the natural frequencies of systems, 401.
- Allan (G. E.), velocity of sound in some binary alloys, 165.
- Arc spectrum of tin, hyperfine structure in the, 315.
- Axially symmetrical stress in a thick tube, 336.
- Banerjee (D. P.) on some expansions containing Laguerre polynomials, and their expressions in terms of Whittaker's confluent hypergeometric functions, 84.
- Bickley (W. G.), experiments in approximating to solutions of a partial differential equation, 50.
- Bikerman (J. J.), friction and adhesion, 67.
- Binney (A. M.), axially symmetrical stress in a thick tube, 336.
- Biot-Savart law and Newton's third law of motion, 171.
- Books, new :—Prof. A. A. Albert, *Introduction to Algebraic Theories*, 86 ; L. Pontrjagin, *Topological Groups*, 86 ; Prof. W. V. D. Hodge, *The Theory and Applications of Harmonic Integrals*, 86 ; The Physical Society ; *Reports on Progress in Physics*.—Vol. VII. 87 ; Dr. A. B. Wood, *A Text-book of Sound*, 88 ; Prof. R. H. Baker, *An Introduction to Astronomy*, 88 ; Prof. S. Chapman and Dr. J. Bartels, *Geomagnetism*, 262 ; Dr. F. Seitz, *The Modern Theory of Solids*, 262 ; Dr. M. Ruhemann, *The Separation of Gases*, 263 ; Prof. J. R. Partington, *Chemical Thermodynamics*, 263 ; H. G. Forder, *The Calculus of Extension*, 263 ; Prof. G. L. Clark, *Applied X-rays*, 264 ; Dr. R. W. B. Pearse and Dr. A. G. Gaydon, *The Identification of Molecular Spectra*, 439 ; E. H. S. van Someren, *Spectrochemical Abstracts*, 440.
- Burington (R. S.) and Cooper (M. D.) on transient similarity and equivalence in n-mesh linear networks, 302.
- Campbell (N. R.) and Francis (V. J.), the shot effect in a saturated diode, 239, 496.
- Chalmers (J. A.), cloud and earth lightning flashes, 77.
- Cherry (E. C.) and Rivlin (R. S.), non-linear distortion, with particular reference to the theory of frequency modulated waves.—Part I., 265
- Cloud and earth lightning flashes, 77.
- Comma in music, 259.
- Conduction of heat in a medium generating heat, 384.
- Cooper (M. D.) and Burington (R. S.), on transient similarity and equivalence in n-mesh linear networks, 302.
- Crystal chemistry, on the presentation of, 106.
- Derivation of the equations of equilibrium of a thin plate, 365.
- Detailed analysis of the gold L spectrum, 1.

- Determination of the shearing stresses in axially symmetrical shafts under torsion by finite difference methods, 33.
- Differential equations of the second order, on the solution of certain, 155.
- satisfying boundary conditions, on the solution of, 483.
- Dufton (A. F.), the comma in music, 259.
- , the world mapped in complementary segments, 436.
- Duncan (W. J.), the admittance method for obtaining the natural frequencies of systems, 401.
- Epistemology and modern physics, 177. (*See also errata*, 352.)
- Erdélyi (A.), note on Heine's integral representation of associated Legendre functions, 351.
- on algebraic Lamé functions, 348.
- Experiments in approximating to solutions of a partial differential equation, 50.
- Flint (H. T.) and Williams (G.), a method for the measurement of impedance by means of Lecher wires, 489.
- Forster (G. O.) and Tolansky (S.), hyperfine structure in the arc spectrum of tin, 315.
- Francis (V. J.) and Campbell (N. R.), the shot effect in a saturated diode, 239, 496.
- Friction and adhesion, 67.
- Garrick (F. J.), the solubility relationships of salts, 353.
- Gatewood (B. E.), thermal stress in long cylindrical bodies, 282.
- Glow discharges, a theory of the anode fall in, 417.
- Gold L spectrum, a detailed analysis of the, 1.
- Heat conduction in composite circular cylinders, 324.
- Heine's integral representation of associated Legendre functions, note on, 351.
- Hodograph, some properties of the, and their application in the construction of the hodograph in the problem of external ballistics, 427.
- Hyperfine structure in the arc spectrum of tin, 315.
- Impedance by means of Lecher wires, a method for the measurement of, 489.
- Investigation concerning G. I. Taylor's correlation coefficient of turbulence, 471.
- Jaeger (J. C.), heat conduction in composite circular cylinders, 324.
- Jeffreys (H.), a derivation of the equations of equilibrium of a thin plate, 365.
- , epistemology and modern physics, 177. (*See also errata*, 352.)
- Kastner (L. J.), the propagation of pressure fluctuations of large amplitude in air columns, 206.
- Kimball (W. S.), partial derivatives of derivatives, 137.
- Kompfner (R.) on a method of correcting the spherical error of electric lenses, especially of those employed with electron microscopes, 410.
- Lenihan (J. M. A.), the shape of a liquid jet, 393.
- Lewis (T.), some properties of the hodograph and their application in the construction of the hodograph in the problem of external ballistics, 427.
- Liquid jet, the shape of a, 393.
- Liquids and liquid mixtures, on a relation between viscosity and density of, 253.
- Magnetization in crystalline media, 225.
- Mathematical tables, project for the computation of, 438.
- Mathur (S. B. L.), Biot-Savart law and Newton's third law of motion, 171.
- Method for the measurement of impedance by means of Lecher wires, 489.

- Morgans (W. R.) on the solution of second order differential equations satisfying boundary conditions, 483.
- Natural frequencies of systems, the admittance method for obtaining the, 401.
- Newing (S. T.), determination of the shearing stresses in axially symmetrical shafts under torsion by finite difference methods, 33.
- Nissan (A. H.), viscosity-temperature function of liquids, 441.
- Non-linear distortion, with particular reference to the theory of frequency modulated waves:—Part I., 265.
- Note on Heine's integral representation of associated Legendre functions, 351.
- Numerical solution of a type of equation, 505.
- On acoustically effective vortex motion in gaseous jets, 245.
- algebraic Lamé functions, 348.
- a method of correcting the spherical error of electron lenses, especially of those employed with electron microscopes, 410.
- a relation between viscosity and density of liquids and liquid mixtures, 253.
- some expansions containing Laguerre polynomials, and their expressions in terms of Whittaker's confluent hypergeometric functions, 84.
- the numerical solution of linear simultaneous equations by an iterative method, 369.
- the presentation of crystal chemistry, 106.
- the solution of certain differential equations of the second order, 155.
- — of second order differential equations satisfying boundary conditions, 483.
- transient similarity and equivalence in n -mesh linear networks, 302.
- Partial derivatives of derivatives, 137.
- Paterson (S.), the conduction of heat in a medium generating heat, 384.
- Peddie (W.), magnetization in crystalline media, 225.
- Project for the computation of mathematical tables, 438.
- Propagation of pressure fluctuations of large amplitude in air columns, 206.
- Plummer (H. C.), the numerical solution of a type of equation, 505.
- Rivlin (R. S.) and Cherry (E. C.), non-linear distortion, with particular reference to the theory of frequency modulated waves.—Part I., 265.
- Roth (L.) on the solution of certain differential equations of the second order, 155.
- Savic (P.) on acoustically effective vortex motion in gaseous jets, 245.
- Schmidt (R. J.) on the numerical solution of linear simultaneous equations by an iterative method, 369.
- Schofield (F. H.), the steady flow of heat from certain objects buried under flat air-cooled surfaces, 437.
- Schumann (T. E. W.), an investigation concerning G. I. Taylor's correlation coefficient of turbulence, 471.
- Shape of a liquid jet, 393.
- Shearing stresses in axially symmetrical shafts under torsion by finite difference methods, determination of the, 33.
- Shot effect in a saturated diode, 239, 496.
- Solubility relationships of salts, the, 353.
- Some properties of the hodograph and their application in the construction of the hodograph in the problem of external ballistics, 427.
- Spherical error of electron lenses, especially of those employed with electron microscopes, on a method of correcting, 410.

- Srinivasan (M. K.) on a relation between viscosity and density of liquid and liquid mixtures, 253.
- Steady flow of a viscous fluid through a leaky tube, 457.
- of heat from certain objects buried under flat air-cooled surfaces, 437.
- Study of the growth of chemical compounds on a copper single crystal by the method of electron diffraction, 89.
- Theory of the anode fall in glow discharges, 417.
- Thermal stress in long cylindrical bodies, 282.
- Thin plate, a derivation of the equations of equilibrium of, 365.
- Tolansky (S.) and Forester (G. O.), hyperfine structure in the arc spectrum of tin, 315.
- Turbulence, an investigation concerning G. I. Taylor's correlation coefficient of, 471.
- Turton (F. J.), steady flow of a viscous fluid through a leaky tube, 457.
- Usmani (I. H.), a study of the growth of chemical compounds on a copper single crystal by the method of electron diffraction, 89.
- Velocity of sound in some binary alloys, 165.
- Viscosity-temperature function of liquids, 441.
- Viscous fluid through a leaky tube, steady flow of a, 457.
- von Engel (A.), a theory of the anode fall in glow discharges, 417.
- Wells (A. F.) on the presentation of crystal chemistry, 106.
- Williams (G.) and Flint (H. T.), a method for the measurement of impedance by means of Lecher wires, 489.
- Woodall (E. A.), a detailed analysis of the gold L spectrum, 1.
- World mapped in complementary segments, 436.

END OF THE THIRTY-SECOND VOLUME

DATE DUE

DEMCO 38-297